Inference in inhomogeneous hidden Markov models with application to ion channel data

Dissertation zur Erlangung

des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm

"PhD School of Mathematical Science (SMS)"

der Georg-August University School of Science (GAUSS)

vorgelegt von

Manuel Diehn

aus Dambeck

Göttingen, 2017

Betreuungsauschuss:

Prof. Dr. Axel Munk
Institut für Mathematische Stochastik, Universität Göttingen

J. Prof. Dr. Daniel Rudolf
Institut für Mathematische Stochastik, Universität Göttingen

Mitglieder der Prüfungskomission:

Referent:
Prof. Dr. Axel Munk
Institut für Mathematische Stochastik, Universität Göttingen

Korreferent:
J. Prof. Dr. Daniel Rudolf
Institut für Mathematische Stochastik, Universität Göttingen

Weitere Mitglieder der Prüfungskomission:

Prof. Dr. Anita Schöbel
Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Prof. Dr. Stephan Waack,
Institut für Informatik, Universität Göttingen

Prof. Dr. Claudia Steinem
Institut für Organische und Biomolekulare Chemie, Universität Göttingen

PD Dr. Timo Aspelmeier
Institut für Mathematische Stochastik, Universität Göttingen

Tag der mündlichen Prüfung: 01.11.2017

Acknowledgment

Preface

Hidden Markov models provide a powerful class of regression models in situations, where the dynamics of a Markov process cannot be observed directly. Its area of applications range from economics, over weather forecasting to biophysiological experiments. One particular example is the study of ion channel gating under a constant environment. In the parametric case, the estimation of the involved parameters is a classical problem in time series analysis and widely investigated.

Ion channel recordings under a changing environment are hardly analyzed and are the main cause for the new model class we introduce. This thesis mainly concerns hidden Markov models with a homogeneous hidden Markov chain and an inhomogeneous observation law, varying in time, but converging to a distribution. The main contribution of this thesis concerns the asymptotic behavior of a quasi-maximum likelihood estimator. In particular, strong consistency and asymptotic normality of this estimator are proven. To this end, we combine asymptotic results of maximum likelihood estimation in homogeneous hidden Markov models with ergodic theory in asymptotic mean stationary processes. The quasi-maximum likelihood estimator is obtained by maximizing the likelihood of the homogeneous process, which can be seen as the limiting process of the observations. It is remarkable that the estimator is computed without any knowledge of the inhomogeneity of the observation law. Therefore, the estimator can be computed straightforward. The model and general methodology are described in Section 2. There we also state the main results of this thesis concerning consistency and asymptotic normality of the quasi-maximum likelihood estimator. Applications of our results can be found in Section 3. We apply the results to a Poisson and a linear Gaussian model. The main steps of the proofs are given in Section 4, whereas technical proofs can be found in the Appendix A. In Section 5 we describe the implementation of likelihood based estimators in hidden Markov models. Especially, we treat the case, when the data is filtered. Simulations and application to ion channel recordings can be found in Section 6. We show statistically significant differences for the interaction of the antibiotic ampicillin with the wild type and with the mutant G103K of the outer membrane channel PorB. These results improve the understanding of potential sources for bacterial resistance and might help to develop new drugs against it to alleviate the severe consequences of multidrug-resistant bacteria.

Contents

List of Symbols

$\mathbb{N}$	set of positive integers
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of positive real numbers
$\lvert S \rvert$	number of elements of a set S
$\mathbb{R}^d$	d-dimensional vector space of real numbers, $d \in \mathbb{N}$
$\mathbb{R}^{d \times d}$	space of $d \times d$ matrices with rea-valued entries, $d \in \mathbb{N}$
Θ	parameter space, $\Theta \subset \mathbb{R}^d, d \in \mathbb{N}$
$\mathscr{B}(G)$	Borel σ-algebra of G
$(G, \mathscr{G})$	measurable space, where G is a set and $\mathscr{G}$ is a σ-algebra
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space, where Ω is a set, $\mathcal{F}$ is a σ-algebra and $\mathbb{P}$ is a probability measure
$\mathbb{P}_\theta$	probability measure, determined by a parameter $\theta \in \Theta$
$\mathbb{E}_\theta$	expected value with respect to $\mathbb{P}_\theta$
Var_θ	covariance matrix with respect to $\mathbb{P}_\theta$
$\mathcal{N}(\mu, \Sigma)$	normal distribution with mean μ and covariance matrix Σ, $\mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d}, d \in \mathbb{N}$
$\mathrm{Poi}(\lambda)$	Poisson distribution with mean λ, $\lambda \in \mathbb{R}_+$
χ_k^2	chi-squared distribution with k degrees of freedom
$B(\theta, \delta)$	Euclidean ball of radius δ centered at θ
$\xrightarrow{\mathbb{P}}$	convergence in probability
$\xrightarrow{\mathcal{D}}$	convergence in distribution
$W \stackrel{\mathcal{D}}{=} V$	two random variables W and V are equal in distribution
$\lvert x \rvert$	Euclidean norm of a vector $x \in \mathbb{R}^d, d \in \mathbb{N}$
$\lVert x \rVert_p$	the ℓ_p-norm of a vector $x \in \mathbb{R}^d, d \in \mathbb{N}, p > 0$
$\lVert A \rVert_p$	the matrix norm induced by the ℓ_p-norm on $\mathbb{R}^d, A \in \mathbb{R}^{d \times d}, d \in \mathbb{N}, p > 0$
$\lambda_{\min}(A)$	the smallest eigenvalue of a semi-positive definite matrix $A \in \mathbb{R}^{d \times d}$
$\mathbb{1}$	indicator function
I	identity matrix
o	a real-valued sequence $(\beta_n)_{n \in \mathbb{N}} = o(\alpha_n)$, where $(\alpha_n)_{n \in \mathbb{N}}$ is another sequence in $\mathbb{R}$, if $\beta_n / \alpha_n \to 0$ as $n \to \infty$.

O a real-valued sequence $(\beta_n)_{n\in\mathbb{N}} = O(\alpha_n)$, where $(\alpha_n)_{n\in\mathbb{N}}$ is another sequence in $\mathbb{R}$, if $\limsup\limits_{n\to\infty} \beta_n/\alpha_n < \infty$.

$O_{\mathbb{P}}$ a sequence of random variables $(X_n)_{n\in\mathbb{N}} = O_{\mathbb{P}}(\alpha_n)$, where $(\alpha_n)_{n\in\mathbb{N}}$ is another sequence in $\mathbb{R}$, if X_n/a_n is bounded in probability

∇ nabla operator

∇^2 Hessian matrix of a real-valued function

λ_n n-times product measure of λ

$\lfloor x \rfloor$ for $x \in \mathbb{R}$ is $\lfloor x \rfloor = \max\{n \in \mathbb{Z} | n \leq x\}$

Introduction

A (homogeneous) hidden Markov model (HMM) is a bivariate stochastic process $(X_n, Y_n)_{n \in \mathbb{N}}$. Here $(X_n)_{n \in \mathbb{N}}$ is a Markov chain with state space S, and $(Y_n)_{n \in \mathbb{N}}$ is, conditioned on $(X_n)_{n \in \mathbb{N}}$, an independent sequence of random variables mapping to a space G, such that the distribution of Y_n depends only on X_n. In a HMM, the Markov chain $(X_n)_{n \in \mathbb{N}}$ is not observable (hidden), but observations of $(Y_n)_{n \in \mathbb{N}}$ are available. HMMs are widely used in different applications of pattern recognition including speech processing, neurophysiology, biology, economy and many more. For readers not familiar with finite state Markov chains, we refer to Appendix B.2 for a short introduction.

In this thesis we model ion channel recordings with a hidden Markov model. Since ion channels are responsible for the flow of ions across cell membranes, it is of particular interest to understand under which circumstances the channel opens and closes. HMMs are with justification well established for analyzing ion channel recordings under stable exogenous conditions, see Ball and Rice (1992), Venkataramanan et al. (2000), Qin et al. (2000), Siekmann et al. (2011) among many others. We stress that for this purpose, there also exist many non-parametric methods, for example Basseville and Benveniste (1983), Colquhoun and Hawkes (1987), Sakmann and Neher (2010), Hotz et al. (2013), Pein et al. (2017b). It is unknown whether the gating behavior of ion channels remains the same if the environment is changing in time, other ion channels do not gate in a stable environment at all, see Yellen (1982), Demo and Yellen (1992), Yellen (1998) and del Camino et al. (2000). In order to stimulate the gating mechanism, experiments with varying voltage have been carried out. Figure 1.1 shows a representative recording of current flow measured under a constantly increasing voltage and a short blockage event of an ion channel. In the case where the applied voltage is linearly increasing Ohm's law suggests that the measured current increases also linearly. Therefore, the quantity of interest is the conductivity of the ion channel. Figure 1.2 shows the conductance level recordings, obtained by dividing the current by the applied voltage.

A natural way to model the ion channel conductance level with a HMM is to assume that the channel attains K states, $K \in \mathbb{N}$. Each state defines whether the channel is closed, open, semi-closed etc. and the corresponding conductance level. Further, it is assumed that the change between the states behaves Markovian. The measurements are a noisy version of each state caused
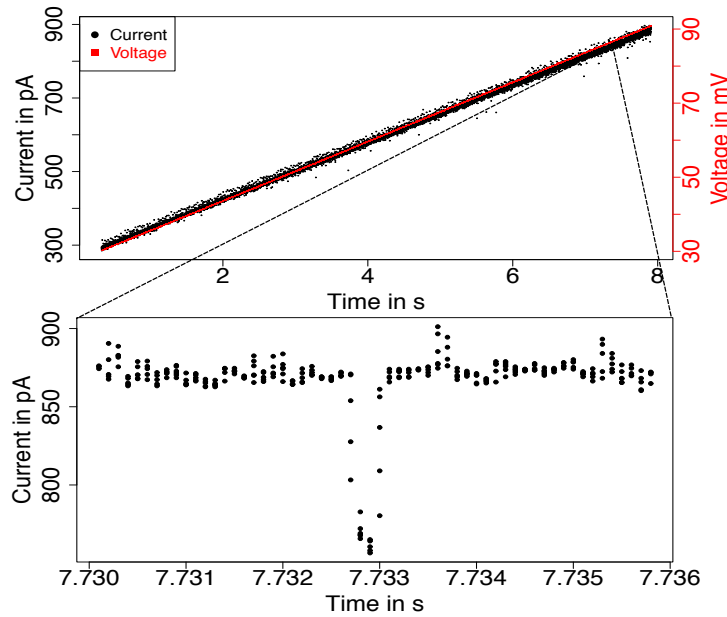
Figure 1.1: Representative current flow of PorB mutant driven by a voltage ramp from $30mV$-$90mV$ (top) and blockage of a PorB mutant protein caused by Ampicillin (bottom).

by errors due to the measuring procedure. From a mathematical point of view the quantities of interest are the corresponding conductance levels, the variance of the noise and the transition rates between the states. In Figure 1.2 it is easily seen that the variance of the measurements changes in time and therefore the conductance levels can not be modeled with a time-homogeneous HMM, rather a time-inhomogeneous modeling seems to be necessary.
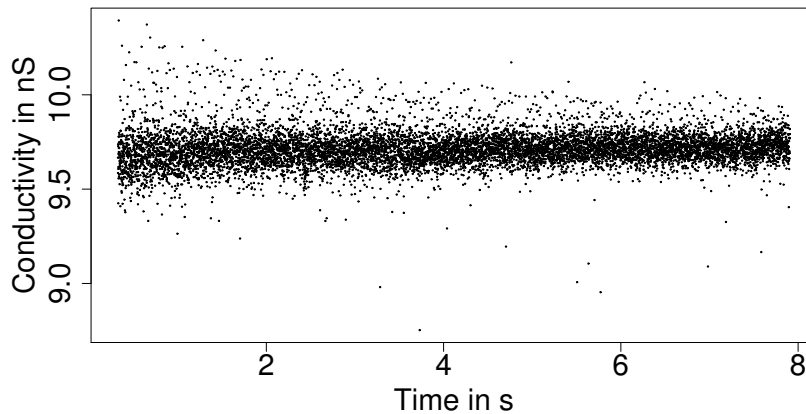
Figure 1.2: Conductivity of a PorB mutant protein. The variance of the measurements decreases in time.

The conditional independence implies that the law of a HMM is determined if the distribution of the underlying Markov chain $(X_n)_{n \in \mathbb{N}}$ and the distribution of Y_n conditioned on X_n are given for all $n \in \mathbb{N}$. In parametric HMMs these distributions are determined by a parameter $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^d, d \in \mathbb{N}$. In classical HMMs the stochastic process is assumed to be homogeneous,

i.e., the conditional distributions are equal for all n and it is assumed that the observations are driven by the unknown "true" parameter $\theta^* \in \Theta$.

The problem of parameter estimation in HMMs has a long history in statistics and related fields, dating back to the 1960's, see Baum and Petrie (1966) and Baum and Eagon (1967). For a profound introduction we refer the reader to the books of Cappé et al. (2007), Zucchini and Macdonald (2009) and Elliott et al. (2008).

In contrast to the classical setting, we consider an *inhomogeneous HMM*, namely a bivariate stochastic process $(X_n, Z_n)_{n\in\mathbb{N}}$, where conditioned on $(X_n)_{n\in\mathbb{N}}$ it is assumed that $(Z_n)_{n\in\mathbb{N}}$ is a sequence of independent random variables on the space G, such that the distribution of Z_n depends not only on the value of X_n, but also changes in n. The additional dependence on n implies that the Markov chain $(X_n, Z_n)_{n\in\mathbb{N}}$ is inhomogeneous.

This motivates us to introduce an *extended HMM*, a trivariate stochastic process $(X_n, Y_n, Z_n)_{n\in\mathbb{N}}$ with the following properties. The sequence $(X_n, Y_n)_{n\in\mathbb{N}}$ is a homogeneous HMM and $(X_n, Z_n)_{n\in\mathbb{N}}$ is an inhomogeneous HMM, such that, given X_n, the distribution of Z_n is getting "closer" to the distribution of Y_n for increasing n. A crucial point here is that $(Z_n)_{n\in\mathbb{N}}$ is observable whereas $(Y_n)_{n\in\mathbb{N}}$ is not. However, Z_n can be considered as "close" to Y_n.

We illustrate this by modeling the conductance level of ion channel data with varying voltage: Here $S = \{0, 1\}$, $G = \mathbb{R}$, $\mu = (\mu^{(1)}, \mu^{(2)}) \in \mathbb{R}^2$ and $\sigma = (\sigma^{(1)}, \sigma^{(2)}) \in (0, \infty)^2$. Assume that $(V_n)_{n\in\mathbb{N}}$ is a real-valued sequence of iid random variables with $V_1 \sim \mathcal{N}(0, 1)$. Further, let $(\varepsilon_n)_{n\in\mathbb{N}}$ be an independent sequence of random variables with $\varepsilon_n \sim \mathcal{N}(0, \beta_n^2)$, where $(\beta_n^2)_{n\in\mathbb{N}} \subset (0, \infty)$ with $\lim_{n\to\infty} \beta_n^2 = 0$. Define

$$Y_n := \mu^{(X_n)} + \sigma^{(X_n)} V_n,$$
$$Z_n := Y_n + \varepsilon_n,$$

where $(Z_n)_{n\in\mathbb{N}}$ is considered as the observations of the channel's conductivity. This extended HMM describes the observed conductance level of ion channel recordings with linearly increasing voltage. Intuitively, here one can already see that for sufficiently large n the influence of ε_n should be negligible and observations of Z_n are "close" to Y_n. Unfortunately none of the theoretic justifications provided in the homogeneous HMM setting are applicable because of the inhomogeneous nature of the noise.

1.1   Main results

The main results of this thesis concern asymptotic properties of the maximum likelihood estimator (MLE) in the described model. Assume that we have a parametrized extended HMM with compact parameter space $\Theta \subseteq \mathbb{R}^d$. For $\theta \in \Theta$ let q_θ^ν be the likelihood function of the homogeneous HMM and p_θ^ν be the likelihood function of the inhomogeneous HMM. Here ν is the initial distribution of the underlying Markov chain. Given observations $z_1, \ldots, z_n$ of $Z_1, \ldots, Z_n$ the goal is to estimate

the "true" parameter $\theta^* \in \Theta$. The maximum likelihood estimator $\theta_{\nu,n}^{\mathrm{ML}}$, given by

$$\theta_{\nu,n}^{\mathrm{ML}} = \underset{\theta \in \Theta}{\mathrm{argmax}} \log p_\theta^\nu(z_1, \ldots, z_n),$$

is the canonical estimator for approaching this problem for the homogeneous case, see Baum and Petrie (1966), Leroux (1992), Douc et al. (2004), Douc et al. (2011). However, the computation of $\theta_{\nu,n}^{\mathrm{ML}}$ requires specific knowledge of the inhomogeneity, in particular of the time-dependent component of the noise. That is the reason for us to introduce a quasi-maximum likelihood estimator, given by

$$\theta_{\nu,n}^{\mathrm{QML}} = \underset{\theta \in \Theta}{\mathrm{argmax}} \log q_\theta^\nu(z_1, \ldots, z_n).$$

This is not a maximum likelihood estimator, since the observations are generated from the inhomogeneous model, whereas q_θ^ν is the likelihood function of the homogeneous model.

Roughly, we assume the following:

- The transition matrix of the hidden finite state space Markov chain is irreducible and aperiodic and satisfies a continuity condition w.r.t. the parameters (see (P1) and (P1')).

- The observable and non-observable random variables $(Z_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ are "close" to each other in a suitable sense (see (C1) – (C3)).

- The homogeneous HMM is well behaving, such that observations of $(Y_n)_{n\in\mathbb{N}}$ would lead to a consistent estimator (see (H1) – (H4)).

- The conditional density of Y_n given X_n is continuously differentiable and integrable w.r.t. to the density of Z_n given X_n (see (CLT1) and (UC1)).

In particular, under the suitable closeness of Z_n to Y_n the estimator $\theta_{\nu,n}^{\mathrm{QML}}$ provides, at least intuitively, a reasonable way for approximating the true parameter θ^*. If the model satisfies the conditions, stated precisely in Section 2.2, then Theorem 2.6, states that almost surely

$$\theta_{\nu,n}^{\mathrm{QML}} \to \theta^*$$

as $n \to \infty$. Hence, the quasi-maximum likelihood estimator is consistent. As a consequence we obtain under an additional assumption that also $\theta_{\nu,n}^{\mathrm{ML}} \to \theta^*$ almost surely as $n \to \infty$.

The asymptotic normality of $\theta_{\nu,n}^{\mathrm{ML}}$ is an application of Theorem 1 in Jensen (2011a) and stated in Corollary 2.11. Additionally, we find that $\theta_{\nu,n}^{\mathrm{QML}}$ is asymptotically normally distributed, see Theorem 2.12. This theorem requires the additional condition (2.16), which in general is difficult to verify.

1.2   Related work

Maximum likelihood estimation in classical HMMs and related model classes has a long history in statistics and goes back to Baum and Petrie (1966) and the extensions in Baum and Eagon (1967) and Baum et al. (1970). These authors considered finite state spaces for the Markov chain and finite observation spaces as well. They proved strong consistency of the MLE under the additional assumption that all transition probabilities are greater than zero. Leroux (1992) generalized the observation state spaces and relaxed the assumption on the transition matrix for the Markov chain to irreducibility. These consistency results uses ergodic theory for stationary processes which is not applicable in our setting since the process we observe is not stationary.

For the first time asymptotic normality of the MLE was addressed by Bickel et al. (1998) who put again the positivity assumption on the transition matrix. Asymptotic properties in more general HMMs have subsequently been investigated in a series of contributions, see Gland and Mevel (2000a), Gland and Mevel (2000b), Douc and Matias (2001), Douc et al. (2004) and Genon-Catalot and Laredo (2006). They used similar ideas and assumed rather restrictive assumptions. The principal idea in proving asymptotic normality uses a central limit theorem for martingales, which is not applicable in the inhomogeneous case.

A breakthrough was achieved by Douc et al. (2011) who proved strong consistency of the MLE for HMMs with general state spaces for the underlying Markov chain. They used the concept of exponential separability to prove directly that the entropy for any $\theta \nsim \theta^*$, even the supremum of a closed ball around θ, is strictly smaller than the entropy of θ^*. The equivalence relation $\sim$ on Θ is introduced in Section 2. We will use some of their results for our consistency proof. However, we work with an inhomogeneous model. We stress that the consistency result of Douc et al. (2011) hold for more general state spaces than our consistency result.

There is some literature which studies asymptotic properties of maximum likelihood estimation of inhomogeneous HMMs, see Ailliot and Pene (2013), Pouzo et al. (2016) and Jensen (2011a). Note that in the setting of homogeneous HMMs the transition probabilities as well as the emission probabilities do not vary over time. In Ailliot and Pene (2013) and Pouzo et al. (2016) asymptotic properties of the maximum likelihood estimator in inhomogeneous Markov switching models are considered. Here the transition probabilities are also influenced by the observations, but the inhomogeneity is different from the time-dependent inhomogeneity considered in our work.

Jensen (2011a) considered the asymptotic normality of M-estimators in the case where the transition probabilities and the emission probabilities vary over time, which is more general than our setting. We apply his result to prove the asymptotic normality of the MLE. However, the quasi-MLE does not satisfy the assumptions stated, but we will use his ideas to show the asymptotic normality of $\theta_{v,n}^{\mathrm{QML}}$. To this end, we introduce the additional condition (2.16) that ensures that the limiting distribution is centered. We stress that, as far as we know, there are no asymptotic results available, if the inhomogenity cannot be modeled.

1.3   Ion channel recordings

The spread of multidrug-resistant bacteria threatens modern medical treatment for infectious diseases causing a large number of fatalities in hospitals. To be able to develop new agents that can combat bacterial infections, the mechanism that contributes to drug resistance needs to be understood. An effective strategy used by Gram-negative bacteria to evade drug treatment is to inhibit the access of antibiotics across the outer membrane, see Delcour (2009). For the influx of antibiotics and other hydrophilic substances through the outer bacterial membrane, ion channels play an important role. They act as filters and select charges and size for a certain range of substrate, see Delcour (2003) and Tanabe et al. (2010).

Ion channels are pore-forming membrane proteins that allow ions to pass through the channel pore. They are present in the membranes of all cells and control the flow of ions across secretory and epithelial cells. They have a significant meaning in the regulation of the osmotic activity and acid-base balance as well as in the saltatory conduction in nerve and muscle cells. For a detailed introduction, we refer to the books of Hille (2001) and Triggle (2006).

The investigation of proteins in artificial membrane systems allows to determine and vary the composition of lipids and proteins and external conditions depending on the biophysical interest. The investigation of electrical properties of cells goes back to first voltage clamp experiments by Cole (1949). Further development of those techniques in Sakmann and Neher (1984) resulting in the so called patch clamp technique enables the scientist to measure the conductivity of isolated ion channels. In 1991, Neher and Sakmann were awarded the Nobel Prize for this work. Very roughly described, a single ion channel is inserted in the (often artificial) membrane surrounded by an electrolyte with an electrode to measure the current while a constant voltage is applied. Figure 1.3 shows a schematic patch clamp configuration. For a more detailed explanation of its various configurations see Sakmann and Neher (2010) and the references therein.

In this thesis, we analyze recordings of the porin PorB of *Neisseria meningitidis* (Nme) performed in the Steinem lab (Institute of Organic and Biomolecular Chemistry, University of Göttingen). Nme is closely related to *Neisseria gonorrhoeae* (Ngo), which is resistant to penicillin and tetracycline. The patch clamp measurements were performed using planar black lipid membranes (BLMs), where "black lipid membrane" refers to the appearance of the prepared planar bilayer. Due to destructive interference of light reflected from both sides of this few nanometer thin bilayer, the membrane appears black. Physical properties such as membrane resistance or membrane capacity can be observed. For a detailed explanation see Winterhalter (2000) or Tien and Ottova (2001). After protein insertion, ampicillin was added from a stock solution (25 mM in 1 M KCl, 10 mM HEPES, pH 7.5 and pH 6.0, respectively) to both sides of the BLM. For control experiments, ampicillin was added only to the trans side. Current traces were recorded at a sampling rate of 50 kHz and filtered with an analogue, four-pole Bessel low-pass filter at 5 kHz.

The very short blockage times and the huge amount of observations and events require an automatic analysis of these recordings with high precision on small temporal scales. In Section 6
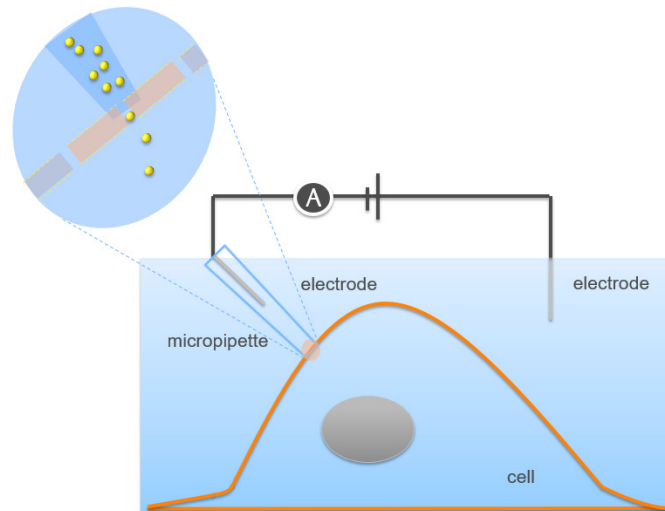
Figure 1.3: Scheme for a patch clamp configuration: a fraction of a membrane is patched by a micropipette and the ion transport across ion channels in the patched membrane part is monitored using two electrodes

we introduce a *forward* algorithm to explore the interaction of the antibiotic ampicillin with the outermembrane porin PorB under constant voltage. We use this algorithm to compute the maximum likelihood estimator under constant voltage and the quasi-maximum likelihood estimator in experiments with varying voltage. Douc and Matias (2001) proved that the maximum likelihood estimator for filtered data is consistent as well. This implies that the transition probability and the dwell-time distributions can be estimated correctly as the number of observations goes to infinity. The asymptotic normality of the maximum likelihood estimator enables us to provide asymptotic confidence intervals for the parameters as well.

We found that the average residence time of ampicillin is statistically significantly longer for the PorB mutant G103K than for the wild type. In conjuncture with other findings this suggests that ampicillin passes the mutant less likely which explains that bacteria with this mutation have an increased resistance against antibiotics. Furthermore, this results match with the results we found for ion channel recordings with varying voltage. Such explorations help to develop new drugs against resistant bacteria.

Assumptions and main results

2.1 Setup and notation

For $K \in \mathbb{N}$ we only consider the case where $S = \{1, \ldots, K\}$ is a finite set and $\mathcal{S}$ denotes the power set of S. Let (G, m) be a Polish space with metric m and corresponding Borel σ-algebra $\mathcal{B}(G)$. The measurable space $(G, \mathcal{B}(G))$ is equipped with a σ-finite reference measure λ. Assume that there is a parametrized family of extended HMMs with compact parameter space $\Theta \subset \mathbb{R}^d$. For each parameter θ the distribution of (X_n, Y_n, Z_n) is specified by

- an initial distribution ν on S and a $K \times K$ transition matrix $P_\theta = (P_\theta(s, t))_{s,t \in S}$ of the Markov chain $(X_n)_{n \in \mathbb{N}}$, such that

$$\mathbb{P}_\theta(X_n = s) = \nu P_\theta^{n-1}(s), \quad s \in S,$$

  where $\nu P_\theta^0 = \nu$ and for $n > 1$,

$$\nu P_\theta^{n-1}(s) = \sum_{s_1, \ldots, s_{n-1} \in S} P_\theta(s_{n-1}, s) \prod_{i=1}^{n-2} P_\theta(s_i, s_{i+1}) \nu(s_1), \quad s \in S;$$

  (Here and elsewhere we use the convention that $\prod_{i=1}^0 a_i = 1$ for any sequence $(a_i)_{i \in \mathbb{N}} \subset \mathbb{R}$.)

- and by the conditional distribution $Q_{\theta,n}$ of (Y_n, Z_n) given $X_n = s$, that is,

$$\mathbb{P}_\theta((Y_n, Z_n) \in C \mid X_n = s) = Q_{\theta,n}(s, C), \qquad C \in \mathcal{B}(G^2)$$

  which satisfies that there are conditional density functions $f_\theta, f_{\theta,n} \colon S \times G \to [0, \infty)$ w.r.t. λ, such that

$$\mathbb{P}_\theta(Y_n \in A \mid X_n = s) = Q_{\theta,n}(s, A \times G) = \int_A f_\theta(s, y) \lambda(\mathrm{d}y), \quad A \in \mathcal{B}(G),$$

$$\mathbb{P}_\theta(Z_n \in B \mid X_n = s) = Q_{\theta,n}(s, G \times B) = \int_B f_{\theta,n}(s, z) \lambda(\mathrm{d}z), \ B \in \mathcal{B}(G).$$

Here the distribution of Y_n given $X_n = s$ is independent of n, whereas the distribution of Z_n

given $X_n = s$ depends through $f_{\theta,n}$ also on n.

We need some further notation and definitions. By $\mathcal{P}(S)$ we denote the set of probability measures on S. To indicate the dependence on the initial distribution, say $v \in \mathcal{P}(S)$, we write $\mathbb{P}^v_\theta$ instead of just $\mathbb{P}_\theta$. To shorten the notation, let $X = (X_n)_{n \in \mathbb{N}}$, $Y = (Y_n)_{n \in \mathbb{N}}$ and $Z = (Z_n)_{n \in \mathbb{N}}$. Further, let $\mathbb{P}^{v,Y}_\theta$ and $\mathbb{P}^{v,Z}_\theta$ be the distributions of Y and Z on $(G^\mathbb{N}, \mathcal{B}(G^\mathbb{N}))$, respectively.

Remark 2.1. *The sequence $(X_n, Y_n)_{n \in \mathbb{N}}$ is a homogeneous Markov chain on $(S \times G, \mathcal{S} \times \mathcal{B}(G))$ with initial distribution*

$$\mathbb{P}^v_\theta((X_1, Y_1) \in C) = \sum_{t \in S} \int_G \mathbb{1}_C(t, y) f_\theta(t, y) \lambda(\mathrm{d}y) v(t), \quad C \in \mathcal{S} \times \mathcal{B}(G),$$

and transition kernel

$$T_\theta((s, y), C) := \sum_{t \in S} \int_G \mathbb{1}_C(t, y') P_\theta(s, t) f_\theta(t, y') \lambda(\mathrm{d}y').$$

In contrast to that, the sequence $(X_n, Z_n)_{n \in \mathbb{N}}$ is an inhomogeneous Markov chain on $(S \times G, \mathcal{S} \times \mathcal{B}(G))$ with initial distribution

$$\mathbb{P}^v_\theta((X_1, Z_1) \in C) = \sum_{t \in S} \int_G \mathbb{1}_C(t, z) f_{\theta,1}(t, z) \lambda(\mathrm{d}z) v(t),$$

and

$$\mathbb{P}^v_\theta((X_n, Z_n) \in C \mid X_{n-1} = s, Z_{n-1} = z) = T_{\theta,n}((s, z), C),$$

with time-dependent transition kernel

$$T_{\theta,n}((s, z), C) := \sum_{t \in S} \int_G \mathbb{1}_C(t, z') P_\theta(s, t) f_{\theta,n}(t, z') \lambda(\mathrm{d}z'), \quad n \geq 2.$$

In our consideration there is a "true" parameter $\theta^* \in \Theta$ and we assume that the transition matrix P_{θ^*} posseses a unique stationary distribution $\pi \in \mathcal{P}(S)$. We have access to a finite length observation of Z. Then, the problem is to find a consistent estimate of θ^* on the basis of the observations without observing $(X_n, Y_n)_{n \in \mathbb{N}}$. Consistency of the estimator of θ^* is limited up to equivalence classes in the following sense. Two parameters $\theta_1, \theta_2 \in \Theta$ are equivalent, written as $\theta_1 \sim \theta_2$, iff there exist two stationary distributions $\mu_1, \mu_2 \in \mathcal{P}(S)$ for $P_{\theta_1}, P_{\theta_2}$, respectively, such that $\mathbb{P}^{\mu_1,Y}_{\theta_1} = \mathbb{P}^{\mu_2,Y}_{\theta_2}$. We illustrate the equivalence relation in the following example.

Example 2.2. Let $(Y_n)_{n \in \mathbb{N}}$ be a sequence of independent, identically distributed random variables, which is also independent of the underlying Markov chain $(X_n)_{n \in \mathbb{N}}$. Then, any two parameters describing the distribution of Y_1 identically are equivalent, although they might lead to a different distribution of $(X_n)_{n \in \mathbb{N}}$.

For the rest of the work assume that each $\theta \in \Theta$ represents its equivalence class.

For an arbitrary finite measure ν on $(S, \mathcal{S})$, $t \in \mathbb{N}$, $x_{t+1} \in S$ and $z_1, \dots, z_t \in G$ define

$$p_\theta^\nu(x_{t+1}; z_1, \dots, z_t) := \sum_{x_1, \dots, x_t \in S} \nu(x_1) \prod_{i=1}^t f_{\theta,i}(x_i, z_i) P_\theta(x_i, x_{i+1}),$$

$$p_\theta^\nu(z_1, \dots, z_t) := \sum_{x_{t+1} \in S} p_\theta^\nu(x_{t+1}; z_1, \dots, z_t).$$

If ν is a probability measure, then $p_\theta^\nu(z_1, \dots, z_n)$ is the likelihood of the observations $(Z_1, \dots, Z_n) = (z_1, \dots, z_n) \in G^n$ for the inhomogeneous HMM $(X_n, Z_n)_{n \in \mathbb{N}}$ with parameter $\theta \in \Theta$ and $X_1 \sim \nu$. Although there are no observations of Y available, we define similar quantities for $(Y_1, \dots, Y_n) = (y_1, \dots, y_n) \in G^n$ by

$$q_\theta^\nu(x_{t+1}, y_1, \dots, y_t) := \sum_{x_1, \dots, x_t \in S} \nu(x_1) \prod_{i=1}^t f_\theta(x_i, y_i) P_\theta(x_i, x_{i+1}),$$

$$q_\theta^\nu(y_1, \dots, y_t) := \sum_{x_{t+1} \in S} q_\theta^\nu(x_{t+1}, y_1, \dots, y_t).$$

Assume for a moment that observations $y_1, \dots, y_n$ of $Y_1, \dots, Y_n$ are available. Then the log-likelihood function of q_θ^ν, with initial distribution $\nu \in \mathcal{P}(S)$, is given by

$$\log q_\theta^\nu(y_1, \dots, y_n)$$

and one can easily consider the maximum likelihood estimator for θ^*. In our setting we do not have access to observations of Y, but have access to observations $z_1, \dots, z_n$ of $Z_1, \dots, Z_n$. We take this trajectory of observations and define a quasi-log-likelihood function

$$\ell_{\nu,n}^Q(\theta) := \log q_\theta^\nu(z_1, \dots, z_n).$$

Now, we approximate θ^* by a quasi-maximum likelihood estimator $\theta_{\nu,n}^{QML}$, that is,

$$\theta_{\nu,n}^{QML} := \operatorname*{argmax}_{\theta \in \Theta} \ell_{\nu,n}^Q(\theta). \tag{2.1}$$

On the other hand, we are interested on the maximum likelihood estimator of a realization $z_1, \dots, z_n$ of $Z_1, \dots, Z_n$. For this define the log-likelihood function

$$\ell_{\nu,n}(\theta) := \log p_\theta^\nu(z_1, \dots, z_n),$$

which leads to the maximum likelihood estimator $\theta_{\nu,n}^{ML}$ given by

$$\theta_{\nu,n}^{ML} := \operatorname*{argmax}_{\theta \in \Theta} \ell_{\nu,n}(\theta). \tag{2.2}$$

Definition 2.3. For $\theta \in \Theta$ and $\delta > 0$ let $B(\theta, \delta)$ be the Euclidean ball of radius δ centered at θ. For any $i \in \mathbb{N}$, let $a_i : \Theta \times S \times S \times G \to \mathbb{R}$ be a function. We say that the sequence $(a_i)_{i \in \mathbb{N}}$ belongs

to the class C_k, $k \in \mathbb{N}$, if there exist constants $\delta_0 > 0$, $K < \infty$, such that for all $i \in \mathbb{N}$ and for all $z \in G$ there exists a function $a_i^0 : G \to \mathbb{R}_+$ with

$$\sup_{s_1, s_2 \in S, \theta \in B(\theta^*, \delta_0)} |a_i(\theta, s_1, s_2, z)| \le a_i^0(z) \quad \text{and} \quad \mathbb{E}_{\theta^*}^{\pi} \left[a_i^0(Z_i)^k \right] \le K.$$

Furthermore, for $k, l \in \mathbb{N}$ the sequence $(a_i)_{i \in \mathbb{N}}$ belongs to the class $C_{k,l}$, if $(a_i)_{i \in \mathbb{N}}$ belongs to C_k and there exist constants $\delta_0 > 0$, $K < \infty$, such that for all $i \in \mathbb{N}$ there exists a function $\bar{a}_i : G \to \mathbb{R}_+$ with

$$\left| a_i(\theta, s_1, s_2, z) - a_i(\theta^*, s_1, s_2, z) \right| \le \left| \theta - \theta^* \right| \bar{a}_i(z) \quad \text{and} \quad \mathbb{E}_{\theta^*}^{\pi} \left[\bar{a}(Z_i)^l \right] \le K$$

for all $\theta \in B(\theta^*, \delta_0)$ and all $s_1, s_2 \in S, z \in G$.

The following notation is used to express the derivatives of $\ell_{\nu,n}^{\mathrm{Q}}$ and $\ell_{\nu,n}$ as sums of conditional expectations. Define the function $\psi : \Theta \times S \times S \times G \to \mathbb{R}^d$, $\psi = (\psi^{(1)}, \dots, \psi^{(d)})$ by

$$\psi^{(r)}(\theta, s_1, s_2, z) := \frac{\partial}{\partial \theta^{(r)}} \left(\log \left(P_\theta(s_1, s_2) f_\theta(s_2, z) \right) \right), \quad r = 1, \dots, d. \tag{2.3}$$

For $i \in \mathbb{N}$, we define $\psi_i : \Theta \times S \times S \times G \to \mathbb{R}^d$, $\psi_i = (\psi_i^{(1)}, \dots, \psi_i^{(d)})$ by

$$\psi_i^{(r)}(\theta, s_1, s_2, z) := \frac{\partial}{\partial \theta^{(r)}} \left(\log \left(P_\theta(s_1, s_2) f_{\theta,i}(s_2, z) \right) \right), \quad r = 1, \dots, d. \tag{2.4}$$

Let n be an integer and I_1 be a finite set with $|I_1| = m$ and $I_1 = \{i_1, \dots, i_m\} \subset \{1, \dots, n\}$. We say I_1 is *ordered* if for all $l, r \in \mathbb{N}$ with $l < r \le m$ we have $i_l < i_r$. Let $z = (z_1, \dots, z_n) \in G^n$. For a finite and ordered set $I_1 = (i_1, \dots, i_m) \subset \{1, \dots, n\}$ we write $z_{|I_1}$ for the projection of z onto the subset G^m indexed by I_1, i.e.,

$$z_{|I_1} = (z_{i_1}, \dots, z_{i_m}) \in G^m.$$

Similarly, for $s \in S^n$ we define the projection $s_{|I_1}$. Furthermore, for two finite and ordered sets I_1, I_2 with $I_2 \subset I_1 \subset \{1, \dots, n\}$ and $s \in S^{I_2}$ and $z \in G^{I_2}$ we define

$$p_{\theta, I_2}^{\nu, I_1}(s \mid z) := \frac{\displaystyle\int_{\substack{y=(y_1,\dots,y_n) \in G^n: \\ y_{|I_1} = z}} \sum_{\substack{x=(x_1,\dots,x_n) \in S^n: \\ x_{|I_2} = s}} \nu(x_1) f_{\theta,1}(x_1, y_1) \prod_{i=2}^{n} P_\theta(x_{i-1}, x_i) f_{\theta,i}(x_i, y_i) \lambda_n(y)}{\displaystyle\int_{\substack{y=(y_1,\dots,y_n) \in G^n: \\ y_{|I_1} = z}} \sum_{x=(x_1,\dots,x_n) \in S^n} \nu(x_1) f_{\theta,1}(x_1, y_1) \prod_{i=2}^{n} P_\theta(x_{i-1}, x_i) f_{\theta,i}(x_i, y_i) \lambda_n(y)}$$

and

$$q_{\theta, I_2}^{\nu, I_1}(s \mid z) := \frac{\displaystyle\int_{\substack{y=(y_1,\dots,y_n) \in G^n: \\ y_{|I_1} = z}} \sum_{\substack{y=(y_1,\dots,y_n) \in G^n: \\ y_{|I_1} = z}} \nu(x_1) f_\theta(x_1, y_1) \prod_{i=2}^{n} P_\theta(x_{i-1}, x_i) f_\theta(x_i, y_i) \lambda_n(y)}{\displaystyle\int_{\substack{y=(y_1,\dots,y_n) \in G^n: \\ y_{|I_1} = z}} \sum_{x=(x_1,\dots,x_n) \in S^n} \nu(x_1) f_\theta(x_1, y_1) \prod_{i=2}^{n} P_\theta(x_{i-1}, x_i) f_\theta(x_i, y_i) \lambda_n(y)}. \tag{2.5}$$

We write $p_{\theta,a:b}^{\nu,r:i}(s_r, \ldots, s_i \mid z_a, \ldots, z_b)$ for $p_{\theta,\{a,\ldots,b\}}^{\nu,\{r,\ldots,i\}}(s_r, \ldots, s_i \mid z_a, \ldots, z_b)$ for the rest of this thesis. Note that for $i, r, a, b \in \mathbb{N}$ with $b \geq i \geq r \geq a$ the conditional density of $X_r = s_r, \ldots, X_i = s_i$ conditioned on $Z_a = z_a, \ldots, Z_b = z_b$ is given by $p_{\theta,\{a,\ldots,b\}}^{\nu,\{r,\ldots,i\}}(s_r, \ldots, s_i \mid z_a, \ldots, z_b)$.

Finally, we define the estimation sums by

$$S_{n,\mathrm{QML}}(\theta) := \sum_{i=2}^{n} \mathbb{E}_\theta^\nu \left[\psi(\theta, X_{i-1}, X_i, Z_i) \frac{q_{\theta,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_1, \ldots, Z_n)}{p_{\theta,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_1, \ldots, Z_n)} \mid Z_1, \ldots, Z_n \right], \qquad (2.6)$$

and

$$S_{n,\mathrm{ML}}(\theta) := \sum_{i=2}^{n} \mathbb{E}_\theta^\nu \left[\psi_i(\theta, X_{i-1}, X_i, Z_i) \mid Z_1, \ldots, Z_n \right]. \qquad (2.7)$$

A standard argument in hidden Markov models, see Section 4 in Bickel et al. (1998), shows that

$$\nabla \ell_{\nu,n}(\theta) = S_{n,\mathrm{ML}}(\theta) + \mathbb{E}_\theta^\nu \left[\nabla \log \left(\nu(X_1) f_{\theta,1}(X_1, Z_1) \right) \mid Z_1, \ldots, Z_n \right]. \qquad (2.8)$$

and

$$\nabla \ell_{\nu,n}^{\mathrm{Q}}(\theta) = S_{n,\mathrm{QML}}(\theta) + \mathbb{E}_\theta^\nu \left[\nabla \log \left(\nu(X_1) f_\theta(X_1, Z_1) \right) \frac{q_{\theta,1:1}^{\nu,1:n}(X_1 \mid Z_1, \ldots, Z_n)}{p_{\theta,1:1}^{\nu,1:n}(X_1 \mid Z_1, \ldots, Z_n)} \mid Z_1, \ldots, Z_n \right]. \qquad (2.9)$$

2.2 Structural conditions for the consistency result

We prove consistency of the quasi-maximum likelihood estimator $\theta_{\nu,n}^{\mathrm{QML}}$ and the maximum likelihood estimator $\theta_{\nu,n}^{\mathrm{ML}}$ under a number of structural assumptions:

Irreducibility and continuity of X

(P1) The transition matrix P_{θ^*} is irreducible.

(P2) The mapping $\theta \mapsto P_\theta$ is continuous w.r.t. some metric induced by a matrix norm.

Closeness of Y and Z

(C1) There exists a number $p > 1$ such that for any $s \in S$ and $\varepsilon > 0$ we have

$$\mathbb{P}_{\theta^*}\left(m(Z_n, Y_n) \geq \varepsilon \mid X_n = s \right) = O(n^{-p}).$$

(C2) There exists an integer $k \in \mathbb{N}$ such that

$$\mathbb{P}_{\theta^*}^\pi \left(\prod_{i=1}^{k-1} \max_{s \in S} \frac{f_{\theta^*,i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} < \infty \right) = 1 \qquad (2.10)$$

$$\mathbb{E}_{\theta^*}^\pi \left[\max_{s' \in S} \frac{f_{\theta^*,n}(s', Z_n)}{f_{\theta^*}(s', Z_n)} \mid X_n = s \right] < \infty, \quad \forall s \in S, n \geq k. \qquad (2.11)$$

and

$$\limsup_{n\to\infty}\left(\mathbb{E}^{\pi}_{\theta^*}\left[\max_{s'\in S}\frac{f_{\theta^*,n}(s',Z_n)}{f_{\theta^*}(s',Z_n)}\mid X_n=s\right]\right)\le 1,\quad \forall s\in S. \tag{2.12}$$

(C3) For every $\theta\in\Theta$ with $\theta\not\sim\theta^*$, there exists a neighborhood $\mathcal{E}_\theta$ of θ such that there exists an integer $k\in\mathbb{N}$ with

$$\mathbb{P}^{\pi}_{\theta^*}\left(\prod_{i=1}^{k-1}\sup_{\theta'\in\mathcal{E}_\theta}\max_{s\in S}\frac{f_{\theta',i}(s,Z_i)}{f_{\theta'}(s,Z_i)}<\infty\right)=1 \tag{2.13}$$

$$\mathbb{E}^{\pi}_{\theta^*}\left[\sup_{\theta'\in\mathcal{E}_\theta}\max_{s'\in S}\frac{f_{\theta',n}(s',Z_n)}{f_{\theta'}(s',Z_n)}\mid X_n=s\right]<\infty,\quad \forall s\in S, n\ge k. \tag{2.14}$$

and

$$\lim_{n\to\infty}\left(\mathbb{E}^{\pi}_{\theta^*}\left[\sup_{\theta'\in\mathcal{E}_\theta}\max_{s'\in S}\frac{f_{\theta',n}(s',Z_n)}{f_{\theta'}(s',Z_n)}\mid X_n=s\right]\right)=1,\quad \forall s\in S. \tag{2.15}$$

Remark 2.4. *The conditions (C1) and (C2) describe a suitable "closeness" of Z_n and Y_n. We will see that (C1) guarantees that $m(Z_n,Y_n)$ converges $\mathbb{P}_{\theta*}$-a.s. to zero whereas (C2) ensures that the ratio of $p^\gamma_{\theta^*}(z_1,\dots,z_n)$ and $q^\gamma_{\theta^*}(z_1,\dots,z_n)$ does not diverge exponentially or faster. Assumption (C3) ensures that for all $\theta\not\sim\theta^*$ the ratio of $p^\gamma_\theta(z_1,\dots,z_n)$ and $q^\gamma_\theta(z_1,\dots,z_n)$ does not diverge exponentially or faster uniformly in $\mathcal{E}_\theta$.*

Well behaving HMM

It is plausible that we are only able to prove consistency in the case where observations of Y would lead to a consistent estimator of θ^*. To guarantee that this is indeed the case we assume:

(H1) For all $s\in S$ let $\mathbb{E}^{\pi}_{\theta^*}\left[\left|\log f_{\theta^*}(s,Y_1)\right|\right]<\infty$.

(H2) For every $\theta\in\Theta$ with $\theta\not\sim\theta^*$, there exists a neighborhood $\mathcal{U}_\theta$ of θ such that

$$\mathbb{E}^{\pi}_{\theta^*}\left[\sup_{\theta'\in\mathcal{U}_\theta}(\log f_{\theta'}(s,Y_1))^+\right]<\infty\qquad\text{for all }s\in S.$$

(H3) The mapping $\theta\mapsto f_\theta(s,y)$ is continuous for any $s\in S, y\in G$.

(H4) For all $s\in S$ and $n\in\mathbb{N}$ let $\mathbb{E}^{\pi}_{\theta^*}\left[\left|\log f_{\theta^*,n}(s,Z_n)\right|\right]<\infty$.

Remark 2.5. *The conditions (H1) – (H3) coincide with the assumptions in Douc et al. (2011) for finite state models and guarantee that the MLE for θ^* based on observations of Y is consistent. The condition (H4) is an additional regularity assumption in the inhomogeneous setting.*

2.3   Consistency theorem

Under the structural assumptions from above we prove the consistency of the quasi-maximum likelihood estimator (2.1) and the maximum likelihood estimator (2.2).

Theorem 2.6. *Assume that the irreducibility and continuity conditions (P1), (P2), the closeness conditions (C1), (C2) and the well behaving HMM conditions (H1) – (H4) are satisfied. Further, let the initial distribution $\nu \in \mathcal{P}(S)$ be strictly positive if and only if π is strictly positive. Then*

$$\theta_{\nu,n}^{\mathrm{QML}} \to \theta^*, \quad \mathbb{P}_{\theta^*}^{\pi}\text{-}a.s.$$

as $n \to \infty$.

Corollary 2.7. *Assume that the conditions of Theorem 2.6 are satisfied. Further, assume that condition (C3) hold. Let the initial distribution $\nu \in \mathcal{P}(S)$ be strictly positive if and only if π is strictly positive. Then*

$$\theta_{\nu,n}^{\mathrm{ML}} \to \theta^*, \quad \mathbb{P}_{\theta^*}^{\pi}\text{-}a.s.$$

as $n \to \infty$.

2.4   Structural conditions for the asymptotic normality result

Asymptotic normality for *M*-estimators in inhomogeneous hidden Markov models was shown in Jensen (2011a). Therefore the assumptions for $\theta_{\nu,n}^{\mathrm{ML}}$ coincide with the assumptions of Jensen (2011a).

Positivity of P_{θ^*}

(P1') We assume that there exist constants $p_0, \delta_0 > 0$ such that

$$\sup_{\theta \in B(\theta^*, \delta_0)} P_\theta(s_1, s_2) \geq p_0 \quad \forall s_1, s_2 \in S.$$

Remark 2.8. *Assumption (P1') is a classical condition in asymptotic theory in hidden Markov models. It guarantees a strong mixing property for the hidden Markov chain. Therefore, the initial probability distribution does not effect the asymptotic behavior of the MLE and the quasi-MLE. Further, the strong mixing of the underlying Markov chain implies a strong mixing property for the conditional Markov chain $(X_n)_{n\in\mathbb{N}}$, conditioned on the observed process $(Z_n)_{n\in\mathbb{N}}$ (see Lemma 4.12).*

Central limit theorem condition

(CLT1) For $r = 1, \ldots, d$ we assume that

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \mathbb{E}_{\theta^*}^{\pi} \left[S_{n,\mathrm{QML}}^{(r)}(\theta^*) \right] = 0 \tag{2.16}$$

and that the constant function sequence $(\psi^{(r)})_{i \in \mathbb{N}}$ belongs to the class C_3. Furthermore, we assume that there exists a constant $c_0 > 0$ and an integer n_0 such that for any $n \geq n_0, n \in \mathbb{N}$, we have

$$\lambda_{\min} \left(n^{-1} \mathrm{Var}_{\theta^*}^{\pi}(S_{n,\mathrm{QML}}(\theta^*)) \right) \geq c_0.$$

Recall that ψ and $S_{n,\mathrm{QML}}$ are given in (2.3) and (2.6), respectively.

(CLT2) For $r = 1, \ldots, d$ we assume that the function sequence $(\psi_i^{(r)})_{i \in \mathbb{N}}$ belongs to the class C_3. Furthermore, we assume there exists a constant $c_0 > 0$ and an integer n_0 such that for any $n \geq n_0, n \in \mathbb{N}$, we have

$$\lambda_{\min} \left(n^{-1} \mathrm{Var}_{\theta^*}^{\pi}(S_{n,\mathrm{ML}}(\theta^*)) \right) \geq c_0.$$

Recall that ψ_i and $S_{n,\mathrm{ML}}$ are given in (2.4) and (2.7), respectively.

Remark 2.9. *Assumption (CLT2) coincides with Assumption 1 in Jensen (2011a) and guarantees a central limit theorem for $S_{n,\mathrm{ML}}$. Assumption (CLT1) is in the same spirit, but has the additional condition (2.16). This condition guarantees that the limiting distribution of $S_{n,\mathrm{QML}}$ has mean zero, which is automatically satisfied for $S_{n,\mathrm{ML}}$. In general it is very difficult to verify (2.16). For the case $S = \{s\}$ the condition (2.16) holds if and only if*

$$\mathbb{E}_{\theta^*}^{\pi} \left[\frac{\partial}{\partial (\theta^*)^{(r)}} \log \left(f_{\theta^*}(s, Z_n) \right) \right] = O(n^{-p}) \quad \forall r = 1, \ldots, d,$$

with $p > 1/2$.

Uniform convergence condition

(UC1) For $n \in \mathbb{N}$ we define the Fisher matrix with respect to $q_{\theta^*}^{\nu}$ by

$$F_{n,\mathrm{QML}} := \frac{1}{n} \mathbb{E}_{\theta^*}^{\pi} \left[-\nabla \left(S_{n,\mathrm{QML}}(\theta^*) \right)^T \right]. \tag{2.17}$$

We assume that there exists a constant $c_0 > 0$ and an integer n_0 such that for $n \geq n_0, n \in \mathbb{N}$, we have

$$\lambda_{\min} \left(F_{n,\mathrm{QML}} \right) \geq c_0.$$

Furthermore, for $r, s = 1, \ldots, d$ we assume that the constant function sequence $(\psi_i^{(r)})_{\in \mathbb{N}}$ belongs to the class $C_{4+\delta}$ for some $\delta > 0$ and that $(\partial \psi^{(r)} / \partial \theta^{(s)})_{i \in \mathbb{N}}$ belongs to the class

$C_{3,2}$.

(UC2) For $n \in \mathbb{N}$ we define the Fisher matrix with respect to $p_{\theta^*}^\gamma$ by

$$F_{n,\mathrm{ML}} := \frac{1}{n} \mathbb{E}_{\theta^*}^\pi \left[-\nabla \left(S_{n,\mathrm{ML}}(\theta^*) \right)^T \right].$$

We assume that there exists a constant $c_0 > 0$ and an integer n_0 such that for $n \geq n_0, n \in \mathbb{N}$, we have

$$\lambda_{\min} \left(F_{n,\mathrm{ML}} \right) \geq c_0.$$

Furthermore, for $r, s = 1, \ldots, d$ we assume that the function sequence $(\psi_i^{(r)})_{i \in \mathbb{N}}$ belongs to the class $C_{4+\delta}$ for some $\delta > 0$ and that $(\partial \psi_i^{(r)} / \partial \theta^{(s)})_{i \in \mathbb{N}}$ belongs to the class $C_{3,2}$.

Remark 2.10. *Condition (UC2) slightly differs from Assumption 2 in Jensen (2011a). In Assumption 2 in Jensen (2011a) the authors assumed that $(\psi_i^{(r)})_{i \in \mathbb{N}}$ belongs to the class C_4. We think that the proof of Lemma 5 in Jensen (2011a) is not valid without the additional δ from our assumption. Further, the authors assumed that $(\partial \psi^{(r)} / \partial \theta^{(s)})_{i \in \mathbb{N}}$ belongs to the class $C_{2,1}$. We think the stronger conditions $C_{3,2}$ is needed in the proof of their Lemma 3. Assumption (UC1) is adapted to the quasi-maximum likelihood estimator. These assumptions are used in proving an uniform convergence results for the Fisher information matrices $F_{n,\mathrm{ML}}$ and $F_{n,\mathrm{QML}}$.*

2.5   Asymptotic normality theorem

Under the structural assumption that prove the consistency of the quasi-maximum likelihood estimator (2.1) and the maximum likelihood estimator (2.2) and the conditions (P1'), (CLT1), (CLT2), (UC1) and (UC2) we can prove the asymptotic normality of the estimators.

Proposition 2.11. *Assume that the positivity condition (P1'), the central limit theorem condition (CLT2) and the uniform convergence condition (UC2) are satisfied. Let I be the d-dimensional identity matrix and for $n \in \mathbb{N}$ define $G_{n,\mathrm{ML}} := \frac{1}{n} \mathrm{Var}_{\theta^*}^\pi (S_{n,\mathrm{ML}}(\theta^*))$. Then for any $v \in \mathcal{P}(S)$ we have*

$$\sqrt{n} G_{n,\mathrm{ML}}^{-1/2} F_{n,\mathrm{ML}} (\theta_{v,n}^{\mathrm{ML}} - \theta^*) \xrightarrow{\mathcal{D}} Z,$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0, I)$ and $G_{n,\mathrm{ML}}^{1/2} G_{n,\mathrm{ML}}^{1/2} = G_{n,\mathrm{ML}}$.

Theorem 2.12. *Assume that the positivity and continuity conditions (P1'), (P2), the closeness conditions (C1), (C2) and the well behaving HMM conditions (H1) – (H4), the central limit theorem condition (CLT1) and the uniform convergence condition (UC1) are satisfied. Let I be the d-dimensional identity matrix and for $n \in \mathbb{N}$ define $G_{n,\mathrm{QML}} := \frac{1}{n} \mathrm{Var}_{\theta^*}^\pi (S_{n,\mathrm{QML}}(\theta^*))$. Then for any $v \in \mathcal{P}(S)$*

$$\sqrt{n} G_{n,\mathrm{QML}}^{-1/2} F_{n,\mathrm{QML}} (\theta_{v,n}^{\mathrm{QML}} - \theta^*) \xrightarrow{\mathcal{D}} Z,$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0, I)$ and $G_{n,\mathrm{QML}}^{1/2} G_{n,\mathrm{QML}}^{1/2} = G_{n,\mathrm{QML}}$.

Application

We consider two models where we verify the structural assumptions from Section 2.2 and Section 2.4 . The Poisson model, see Section 3.1, illustrates a simple example with countable observation space. The linear Gaussian model, see Section 3.2, is an extension of the model which describes the conductivity of ion channels. Here we have multiple and possibly correlated observations.

3.1   Poisson model

Let $X = (X_n)_{n \in \mathbb{N}}$ be a finite state Markov chain on $S = \{1, \ldots, K\}$ induced by an irreducible stochastic matrix P_{θ^*} with stationary distribution π. For $i = 1, \ldots, K$ let $\lambda_{\theta^*}^{(i)} > 0$ and define the vector $\lambda_{\theta^*} = (\lambda_{\theta^*}^{(1)}, \ldots, \lambda_{\theta^*}^{(K)})$. For simplicity, we assume that

$$\theta = (P_\theta(1, 1), \ldots, P_\theta(1, K-1), P_\theta(2, 1), \ldots, P_\theta(K-1, K-1), \lambda_\theta^{(1)}, \ldots, \lambda_\theta^{(K)})^T,$$

so $\Theta \subset \mathbb{R}^{(K-1)^2+K}$. Conditioned on X the non-observed homogeneous sequence $Y = (Y_n)_{n \in \mathbb{N}}$ is an independent sequence of Poisson-distributed random variables with parameter $\lambda_{\theta^*}^{(X_n)}$. In other words, given X_n we have $Y_n \sim \text{Poi}(\lambda_{\theta^*}^{(X_n)})$. The observed sequence $Z = (Z_n)_{n \in \mathbb{N}}$ is determined by

$$Z_n = Y_n + \varepsilon_n,$$

where $(\varepsilon_n)_{n \in \mathbb{N}}$ is an independent sequence of random variables with $\varepsilon_n \sim \text{Poi}(\beta_n)$. Here $(\beta_n)_{n \in \mathbb{N}}$ is a sequence of positive real numbers such that there exists a $p > 1$ with

$$\beta_n = O(n^{-p}). \tag{3.1}$$

We also assume that $(\varepsilon_n)_{n \in \mathbb{N}}$ is independent of Y. Note that the observation space is given by $G = \mathbb{N} \cup \{0\}$ equipped with the counting measure denoted by λ.

To obtain consistency of the two maximum likelihood estimators we need to check the conditions (P1), (P2), (C1) – (C3) and (H1) – (H4):

To (P1) and (P2): By the assumptions in this scenario those conditions are satisfied.

To (H1) – (H4): For $\theta \in \Theta$, $s \in S$ and $y \in G$ we have

$$\left| \log f_\theta(s, y) \right| = -\log \left(\frac{\left(\lambda_\theta^{(s)} \right)^y}{y!} \exp(-\lambda_\theta^{(s)}) \right) = -y \log(\lambda_\theta^{(s)}) + \log(y!) + \lambda_\theta^{(s)}$$

$$\leq -y \log(\lambda_\theta^{(s)}) + y^2 + \lambda_\theta^{(s)}.$$

Hence

$$\mathbb{E}_{\theta^*}^\pi \left[\left| \log f_{\theta^*}(s, Y_1) \right| \right] \leq -\log(\lambda_{\theta^*}^{(s)}) \sum_{s=1}^K \pi(s) \lambda_{\theta^*}^{(s)} + \sum_{s=1}^K \pi(s) \left(\left(\lambda_{\theta^*}^{(s)} \right)^2 + \lambda_{\theta^*}^{(s)} \right) + \lambda_{\theta^*}^{(s)} < \infty$$

and (H1) is verified. A similar calculation leads to the fact that (H4) holds. Condition (H2) follows simply by $(\log f_\theta(s, y))^+ = 0$. Condition (H3) follows by the continuity in the parameter of the probability function of the Poisson distribution and the continuity of the mapping $\theta \mapsto (P_\theta, \lambda_\theta)$.

To (C1) – (C3): For any $\delta > 0$ and any $s \in S$ we have

$$\mathbb{P}_\theta^\pi \left(|Z_n - Y_n| \geq \delta \mid X_n = s \right) = \mathbb{P}_\theta^\pi (\varepsilon_n \geq \delta) \leq 1 - \mathbb{P}_\theta^\pi (\varepsilon_n = 0) = 1 - \exp(-\beta_n).$$

For p and C as in (3.1) it follows that

$$\limsup_{n \to \infty} \frac{1 - \exp(-\beta_n)}{n^{-p}} \leq C,$$

which proves (C1). Observe that for any $s \in S, z \in G$ we have

$$\max_{s \in S} \frac{f_{\theta^*, n}(s, z)}{f_{\theta^*}(s, z)} = \max_{s \in S} \frac{\left(\beta_n + \lambda_{\theta^*}^{(s)} \right)^z}{\left(\lambda_{\theta^*}^{(s)} \right)^z} \exp(-\beta_n) = (a_n)^z \exp(-\beta_n),$$

with $a_n = \max_{s \in S} \frac{\beta_n + \lambda_{\theta^*}^{(s)}}{\lambda_{\theta^*}^{(s)}}$. Now we verify (C2) with $k = 1$. We have

$$\mathbb{E}_{\theta^*}^\pi \left[\max_{s' \in S} \frac{f_{\theta^*, n}(s', Z_n)}{f_{\theta^*}(s', Z_n)} \mid X_n = s \right] = \mathbb{E}_{\theta^*}^\pi \left[a_n^{Z_n} \exp(-\beta_n) \mid X_n = s \right]$$

$$= \exp \left(\lambda_{\theta^*}^{(s)} (a_n - 1) - \beta_n \right) < \infty \quad \forall n \in \mathbb{N}, s \in S.$$

Fix $s \in S$, and note that

$$\limsup_{n \to \infty} \left(\mathbb{E}_{\theta^*}^\pi \left[\max_{s' \in S} \frac{f_{\theta^*, n}(s', Z_n)}{f_{\theta^*}(s', Z_n)} \mid X_n = s \right] \right)$$

$$= \limsup_{n \to \infty} \exp \left(\lambda_{\theta^*}^{(s)} (a_n - 1) - \beta_n \right) = 1.$$

The last equality follows by the fact that $\lim_{n \to \infty} a_n = 1$ and $\lim_{n \to \infty} \beta_n = 0$. Condition (C3) follows by similar arguments.

The application of Theorem 2.6 and Corollary 2.7 leads to the following result.

Corollary 3.1. *For any initial distribution $v \in \mathcal{P}(S)$ which is strictly positive if and only if π is strictly positive, we have in the setting of the Poisson model that*

$$\theta_{v,n}^{\mathrm{QML}} \to \theta^*, \quad \mathbb{P}_{\theta^*}^{\pi}\text{-}a.s.$$

and

$$\theta_{v,n}^{\mathrm{ML}} \to \theta^*, \quad \mathbb{P}_{\theta^*}^{\pi}\text{-}a.s.$$

as $n \to \infty$.

In order to apply Proposition 2.11 and Theorem 2.12, we have to make additional assumptions. We assume that P_{θ^*} is positive. Further, we assume that condition (2.16) holds and that there exists a constant c_0 and an integer n_0 such that for all $n \geq n_0, n \in \mathbb{N}$, we have

$$\lambda_{\min}\left(n^{-1}\mathrm{Var}_{\theta^*}^{\pi}(S_{n,\mathrm{QML}}(\theta^*))\right) \geq c_0, \qquad \lambda_{\min}\left(n^{-1}\mathrm{Var}_{\theta^*}^{\pi}(S_{n,\mathrm{ML}}(\theta^*))\right) \geq c_0$$

and

$$\lambda_{\min}\left(F_{n,\mathrm{QML}}\right) \geq c_0, \qquad \lambda_{\min}\left(F_{n,\mathrm{ML}}\right) \geq c_0.$$

Now, we check the conditions (P1'), (CLT1), (CLT2), (UC1) and (UC2):

To (P1'): By the additional assumptions in this scenario this condition is satisfied.

To (CLT1) and (CLT2):

Condition (2.16) is satisfied by assumption. Unfortunately, we cannot verify this condition analytically. Simulations reveal that (2.16) holds, if

$$\beta_n = o(n^{-1/2}).$$

We refer to Section 6 for more details. Recall that

$$\psi(\theta, s_1, s_2, z) = \frac{\partial}{\partial \theta} \log(P_{\theta}(s_1, s_2)f_{\theta}(s_2, z)), \quad \theta \in \Theta, s_1, s_2 \in S, z \in G.$$

Now, fix an integer $r \in \{1, \ldots, d\}$. If $\theta^{(r)} = P_{\theta}(j, k)$ for some $j, k \in S$, then

$$\frac{\partial}{\partial \theta^{(r)}} \log(P_{\theta}(s_1, s_2)f_{\theta}(s_2, z)) = \frac{\mathbb{1}_{\{j\}}(s_1)\mathbb{1}_{\{k\}}(s_2)}{P_{\theta}(j, k)}.$$

Clearly the constant function sequence $\left(\frac{\mathbb{1}_{\{j\}}(s_1)\mathbb{1}_{\{k\}}(s_2)}{P_{\theta}(j,k)}\right)_{i \in \mathbb{N}}$ belongs to the class C_3 by assumption (P1'). If $\theta^{(r)} = \lambda_{\theta}^{(j)}$ for some $j \in S$, we have

$$\frac{\partial}{\partial \theta^{(r)}} \log(P_{\theta}(s_1, s_2)f_{\theta}(s_2, z)) = \left(\frac{z}{\lambda_{\theta}^{(j)}} - 1\right)\mathbb{1}_{\{j\}}(s_2).$$

Since for any $i \in \mathbb{N}$, Z_i is a mixture of Poisson distributed random variables, it follows that $(\psi^{(r)})_{i \in \mathbb{N}}$ belongs to the class C_3.

The last condition of (CLT1) is satisfied by assumption. This condition concerns positive

definiteness and is classical in HMMs. The condition usually is difficult to verify, see Theorem 1 in Bickel et al. (1998). Assumption (CLT2) is satisfied by similar arguments.

To (UC1) and (UC2): Note that the first condition of assumption (UC1) again concerns positive definiteness. This condition is satisfied by assumption.

Similarly as above, one can show for $r, s = 1, \ldots, d$ that $(\psi^{(r)})_{i \in \mathbb{N}}$ belongs to the class C_5 and that $(\partial \psi^{(r)} / \partial \theta^{(s)})_{i \in \mathbb{N}}$ belongs to the class C_3.

Now, we fix two integers $r, s \in \{1, \ldots, d\}$. Assume $\theta^{(r)} = P_\theta(j, k)$ for some $j, k \in S$. It follows that $\partial \psi^{(r)} / \partial \theta^{(s)} = 0$ whenever $s \neq r$. For $s = r$, we have

$$\frac{\partial}{\partial \theta^{(s)}} \psi^{(r)}(\theta, s_1, s_2, z) = \frac{-\mathbb{1}_{\{j\}}(s_1)\mathbb{1}_{\{k\}}(s_2)}{(P_\theta(j, k))^2}.$$

It follows that $\partial \psi^{(r)} / \partial \theta^{(s)}$ belongs to the class $C_{3,2}$ with

$$\bar{\psi}_i(z) = \sup_{\theta \in B(\theta^*, \delta_0)} \frac{2}{(P_\theta(s_1, s_2))^3}.$$

Assume now that $\theta^{(r)} = \lambda_\theta^{(j)}$ for some $j \in S$. It follows that $\partial \psi^{(r)} / \partial \theta^{(s)} = 0$ whenever $s \neq r$. For $s = r$, we have

$$\frac{\partial}{\partial \theta^{(s)}} \psi^{(r)}(\theta, s_1, s_2, z) = \frac{-z \mathbb{1}_{\{j\}}(s_2)}{\left(\lambda_\theta^{(r)}\right)^2}.$$

It follows that $\partial \psi^{(r)} / \partial \theta^{(s)}$ belongs to the class $C_{3,2}$ with

$$\bar{\psi}_i(z) = \sup_{\theta \in B(\theta^*, \delta_0)} \frac{z}{\left(\lambda_\theta^{(r)}\right)^3}.$$

Assumption (UC2) follows by similar arguments.

The application of Theorem 2.12 and Proposition 2.11 leads to the following result.

Corollary 3.2. *For any initial distribution $\nu \in \mathcal{P}(S)$, we have in the setting of the Poisson model that*

$$\sqrt{n} G_{n,\mathrm{QML}}^{-1/2} F_{n,\mathrm{QML}}(\theta_{\nu,n}^{\mathrm{QML}} - \theta^*) \xrightarrow{\mathcal{D}} Z,$$

and

$$\lim_{n \to \infty} \sqrt{n} G_{n,\mathrm{ML}}^{-1/2} F_{n,\mathrm{ML}}(\theta_{\nu,n}^{\mathrm{ML}} - \theta^*) \xrightarrow{\mathcal{D}} Z$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0, I)$, $G_{n,\mathrm{QML}}^{1/2} G_{n,\mathrm{QML}}^{1/2} = G_{n,\mathrm{QML}}$ and $G_{n,\mathrm{ML}}^{1/2} G_{n,\mathrm{ML}}^{1/2} = G_{n,\mathrm{ML}}$.

3.2   Linear Gaussian model

Let $(X_n)_{n \in \mathbb{N}}$ be a finite state Markov chain on $S = \{1, \ldots, K\}$ induced by an irreducible stochastic matrix P_{θ^*} with stationary distribution π. For $i = 1, \ldots, K$ let $\mu_{\theta^*}^{(i)} \in \mathbb{R}^M$, $\Sigma_{\theta^*}^{(i)} \in \mathbb{R}^{M \times M}$ with full rank, where $M \in \mathbb{N}$. We set $\mu_{\theta^*} = (\mu_{\theta^*}^{(1)}, \ldots, \mu_{\theta^*}^{(K)})$ and $\Sigma_{\theta^*} = (\Sigma_{\theta^*}^{(1)}, \ldots, \Sigma_{\theta^*}^{(K)})$. The sequences

$Y = (Y_n)_{n \in \mathbb{N}}$ and $Z = (Z_n)_{n \in \mathbb{N}}$ are defined by

$$Y_n = \mu_{\theta^*}^{(X_n)} + \Sigma_{\theta^*}^{(X_n)} V_n$$
$$Z_n = Y_n + \varepsilon_n.$$

Here $(V_n)_{n \in \mathbb{N}}$ is an iid sequence of random vectors with $V_n \sim \mathcal{N}(0, I)$, where $I \in \mathbb{R}^{M \times M}$ denotes the identity matrix, and $(\varepsilon_n)_{n \in \mathbb{N}}$ is a sequence of independent random vectors with $\varepsilon_n \sim \mathcal{N}(0, \beta_n^2 I)$, where $(\beta_n)_{n \in \mathbb{N}}$ is a sequence of positive real numbers such that there is a number $q > 0$ such that

$$\beta_n = O(n^{-q}). \tag{3.2}$$

For simplicity, we assume that

$$\theta = \left(P_\theta(1,1), \dots, P_\theta(1, K-1), \dots, P_\theta(K-1, K-1), (\mu_\theta^{(1)})^T, \dots, (\mu_\theta^{(K)})^T, \right.$$
$$\left. \Sigma_\theta^{(1)} \left(\Sigma_\theta^{(1)} \right)^T (1,1), \dots, \Sigma_\theta^{(K)} \left(\Sigma_\theta^{(K)} \right)^T (M, M) \right)^T,$$

so $\Theta \subset \mathbb{R}^{(K-1)^2 + MK + M^2 K}$. Furthermore, note that $G = \mathbb{R}^M$ and λ is the M-dimensional Lebesgue measure.

To obtain consistency of the two maximum likelihood estimators we need to check the conditions (P1), (P2), (C1) – (C3) and (H1) – (H4):

To (P1) and (P2): By definition of the model this conditions are satisfied.

To (H1) – (H4): For a matrix $A \in \mathbb{R}^{M \times M}$ denote $A^2 = AA^T$ and $A^{-2} = (A^2)^{-1}$. Note that for $s \in S$, $\theta \in \Theta$ and $y, z \in G$ we have

$$f_\theta(s, y) = \frac{(2\pi)^{-M/2}}{\det\left(\left(\Sigma_\theta^{(s)} \right)^2 \right)^{1/2}} \exp\left(-\frac{1}{2}(y - \mu_\theta^{(s)})^T \left(\Sigma_\theta^{(s)} \right)^{-2} (y - \mu_\theta^{(s)}) \right),$$

$$f_{\theta, n}(s, z) = \frac{(2\pi)^{-M/2}}{\det\left(\left(\Sigma_\theta^{(s)} \right)^2 + \beta_n^2 I \right)^{1/2}} \exp\left(-\frac{1}{2}(z - \mu_\theta^{(s)})^T \left(\left(\Sigma_\theta^{(s)} \right)^2 + \beta_n^2 I \right)^{-1} (z - \mu_\theta^{(s)}) \right).$$

Further, observe that $\det\left(\left(\Sigma_\theta^{(s)} \right)^2 \right) > 0$ for all $s \in S$. For some constant $C_1 > 0$ we have

$$\mathbb{E}_{\theta^*}^\pi \left[\left| \log f_\theta(s, Y_1) \right| \right] \le C_1 + \mathbb{E}_{\theta^*}^\pi \left[\frac{1}{2}(Y_1 - \mu_\theta^{(s)})^T \left(\Sigma_\theta^{(s)} \right)^{-2} (Y_1 - \mu_\theta^{(s)}) \right] < \infty,$$

since for each $i, j \in \{1, \dots, M\}$ we have $\mathbb{E}_{\theta^*}^\pi \left[Y_1^{(i)} Y_1^{(j)} \right] < \infty$ for $Y_1 = (Y_1^{(1)}, \dots, Y_1^{(M)})$. By this estimate (H1) and (H2) follows easily. Condition (H4) follows by similar arguments. More detailed, we have that β_n^2 is finite and converges to zero as well as that there exists a constant

$C_2 > 0$ such that

$$\mathbb{E}^\pi_{\theta^*} \left[\left| \log f_{\theta^*,n}(s, Z_n) \right| \right] \leq C_2 + \mathbb{E}^\pi_{\theta^*} \left[\frac{1}{2} (Z_n - \mu_s)^T \left(\left(\Sigma^{(s)}_\theta \right)^2 + \beta_n^2 I \right)^{-1} (Z_n - \mu_s) \right].$$

For all $n \in \mathbb{N}$ the right-hand side of the previous inequality is finite, since for each $i, j \in \{1, \ldots, M\}$ we have $\mathbb{E}^\pi_{\theta^*}[Z_n^{(i)} Z_n^{(j)}] < \infty$, with $Z_n = (Z_n^{(1)}, \ldots, Z_n^{(M)})$. Finally condition (H3) is satisfied by the continuity of the conditional density and the continuity of the mapping $\theta \mapsto (P_\theta, \mu_\theta, \Sigma_\theta)$.

To (C1) – (C3): Here m is the Euclidean metric in $\mathbb{R}^M$ such that $|\varepsilon_n| = m(Y_n, Z_n)$. Fix some $p > 1$ and observe that for any $\delta > 0$ and $s \in S$ we have

$$\mathbb{P}^\pi_{\theta^*} (m(Y_n, Z_n) > \delta \mid X_n = s) = \mathbb{P}^\pi_{\theta^*} (|\varepsilon_n| > \delta) = \mathbb{P}^\pi_{\theta^*} \left(\beta_n^2 \chi_M^2 > \delta^2 \right) \leq \frac{\mathbb{E}^\pi_{\theta^*} \left[(\chi_M^2)^{p/q} \right] \beta_n^{2p/q}}{\delta^{2p/q}},$$

where χ_M^2 is a chi-squared distributed random variable with M degrees of freedom. By the fact that $\mathbb{E}^\pi_{\theta^*} \left[(\chi_M^2)^{p/q} \right] < \infty$ and (3.2) we obtain that condition (C1) is satisfied with $p > 1$.

The requirement of (2.10) of (C2) holds for any $k \in \mathbb{N}$, since the density of normally distributed random vectors is strictly positive and finite. Observe that

$$\max_{s \in S} \frac{f_{\theta,n}(s, Z_n)}{f_\theta(s, Z_n)} \leq C_n \max_{s \in S} \exp \left(-\frac{1}{2} (Z_n - \mu^{(s)}_\theta)^T \left(((\Sigma^{(s)}_\theta)^2 + \beta_n I)^{-1} - (\Sigma^{(s)}_\theta)^{-2} \right) (Z_n - \mu^{(s)}_\theta) \right),$$

with

$$C_n := \max_{s \in S} \frac{\left(\det \left(\left(\Sigma^{(s)}_\theta \right)^2 \right) \right)^{1/2}}{\left(\det \left(\left(\Sigma^{(s)}_\theta \right)^2 + \beta_n^2 I \right) \right)^{1/2}}.$$

Note that $\lim_{n \to \infty} C_n = 1$. Since for an invertible matrix $A \in \mathbb{R}^{M \times M}$, $A \mapsto A^{-1}$ is continuous and $\Sigma^s_{\theta^*}$ has full rank, it follows that

$$\lim_{n \to \infty} \left(\left(\Sigma^{(s)}_\theta \right)^2 + \beta_n^2 I \right)^{-1} = \left(\Sigma^{(s)}_\theta \right)^{-2}.$$

Set $(\Sigma^{(s)}_\theta)_n^2 := (\Sigma^{(s)}_\theta)^2 + \beta_n^2 I$ and define $B_n = B_{n,s} := (\Sigma^{(s)}_\theta)^{-2} - (\Sigma^{(s)}_\theta)_n^{-2}$. Note that the entries of B_n converge to zero when n goes to infinity. Further, by the fact that $(B_n)_{n \in \mathbb{N}}$ is a sequence of symmetric, positive definite matrices there exist sequences of orthogonal matrices $(U_n)_{n \in \mathbb{N}} \subset \mathbb{R}^{M \times M}$ and diagonal matrices $(D_n)_{n \in N} \subset \mathbb{R}^{M \times M}$ such that

$$B_n = U_n^T D_n^{1/2} D_n^{1/2} U_n.$$

Of course, U_n and D_n depend on s. We define a sequence of random vectors $(W_{n,s})_{n \in \mathbb{N}}$ by setting $W_{n,s} := U_n D_n^{1/2} (Z_n - \mu^{(s)}_\theta)$, such that

$$(Z_n - \mu^{(s)}_\theta)^T \left((\Sigma^{(s)}_\theta)^{-2} - ((\Sigma^{(s)}_\theta)^2 + \beta_n I)^{-1} \right) (Z_n - \mu^{(s)}_\theta)$$
$$= (Z_n - \mu^{(s)}_\theta)^T B_n (Z_n - \mu^{(s)}_\theta) = W_{n,s}^T W_{n,s}.$$

The random variable Z_i conditioned on $X_i = x$ is normally distributed with mean $\mu_\theta^{(x)}$ and covariance matrix $(\Sigma_\theta^{(x)})_n^2$. Hence $W_{i,s}$, conditioned on $X_i = x$, satisfies

$$W_{i,s} \sim \mathcal{N}(\tilde{\mu}_i, A_i),$$

with

$$\tilde{\mu}_i = U_i^T D_i^{1/2} (\mu_\theta^{(x)} - \mu_\theta^{(s)})$$

and

$$A_i = U_i^T D_i^{1/2} (\Sigma_\theta^{(x)})_i^2 (U_i^T D_i^{1/2})^T.$$

Since A_i is symmetric and positive definite, we find sequences of orthogonal matrices $(U_n')_{n \in \mathbb{N}}$ and diagonal matrices $(D_n')_{n \in \mathbb{N}}$ depending on x and s such that

$$A_i = U_i' D_i'^{1/2} D_i'^{1/2} U_i'^T.$$

Let $(N_i)_{i \in \mathbb{N}}$ be an iid sequence random vectors with $N_i \sim \mathcal{N}(0, I)$ and denote $N_i = (N_i^{(1)}, \ldots, N_i^{(M)})$. Then

$$\begin{aligned}
W_{i,s}^T W_{i,s} = |W_{i,s}|^2 &\overset{\mathcal{D}}{=} \left| U_i' D_i'^{1/2} (N_i + D_i'^{-1/2} U_i'^T \tilde{\mu}_i) \right|^2 \\
&= \left| D_i'^{1/2} (N_i + D_i'^{-1/2} U_i'^T \tilde{\mu}_i) \right|^2 = \sum_{j=1}^M D_i'(j, j) \left(N_i^{(j)} + (D_i'^{-1/2} U_i'^T \tilde{\mu}_i)^{(j)} \right)^2.
\end{aligned}$$

For any $t < \min_{j=1,\ldots,M} D_i'(j, j)^{-1}$ the moment generating function of a chi-squared distributed random variable with one degree of freedom and non-centrality parameter $(D_i'^{-1/2} U_i'^T \tilde{\mu}_i)^{(j)}$ is well-defined and we obtain

$$\begin{aligned}
&\mathbb{E}_{\theta^*}^\pi \left[\exp\left(\frac{t}{2} W_{i,s'}^T W_{i,s'} \right) \mid X_i = s \right] \\
&= \prod_{j=1}^M (1 - 2(\tfrac{t}{2}) D_i'(j, j))^{-1/2} \exp\left(\frac{(D_i'^{-1/2} U_i'^T \tilde{\mu}_i)^{(j)} (\tfrac{t}{2}) D_i'(j, j)}{1 - 2(\tfrac{t}{2}) D_i'(j, j)} \right) \\
&= \prod_{j=1}^M (1 - t D_i'(j, j))^{-1/2} \exp\left(\frac{(D_i'^{-1/2} U_i'^T \tilde{\mu}_i)^{(j)} (\tfrac{t}{2}) D_i'(j, j)}{1 - t D_i'(j, j)} \right) \to 1
\end{aligned}$$

as $i \to \infty$, since $\lim_{i \to \infty} D_i'(j, j) = 0$ for all $j = 1, \ldots, M$. We can choose k sufficiently large, such that $K < \min_{j=1,\ldots,M} D_i'(j, j)^{-1}$ for all $i \geq k$. We find that

$$\begin{aligned}
\mathbb{E}_{\theta^*}^\pi \left[\max_{s' \in S} \exp\left(\frac{1}{2} W_{k,s'}^T W_{k,s'} \right) \mid X_k = s \right] &\leq \mathbb{E}_{\theta^*}^\pi \left[\sum_{s' \in S} \exp\left(\frac{1}{2} W_{k,s'}^T W_{k,s'} \right) \mid X_k = s \right] \\
&\leq \prod_{s' \in S} \left(\mathbb{E}_{\theta^*}^\pi \left[\exp\left(\frac{K}{2} W_{k,s'}^T W_{k,s'} \right) \mid X_k = s \right] \right)^{1/K},
\end{aligned}$$

where we used the generalized Hölder inequality in the last estimate. Then, by taking the limit

superior we obtain that the right-hand side of the previous inequality goes to one for $k \to \infty$ such that (C2) holds. Condition (C3) can be verified similarly.

Corollary 3.3. *For any initial distribution $v \in \mathcal{P}(S)$ which is strictly positive if and only if π is strictly positive, we have in the setting of the linear Gaussian model that*

$$\theta_{v,n}^{\mathrm{QML}} \to \theta^*, \quad \mathbb{P}_{\theta^*}^{\pi}\text{-}a.s.$$

and

$$\theta_{v,n}^{\mathrm{ML}} \to \theta^*, \quad \mathbb{P}_{\theta^*}^{\pi}\text{-}a.s.$$

as $n \to \infty$.

In order to apply Proposition 2.11 and Theorem 2.12, we have to make additional assumptions as in the Poisson model. We assume that P_{θ^*} is positive. Further, we assume that condition (2.16) holds and that there exists a constant c_0 and an integer n_0 such that for all $n \geq n_0, n \in \mathbb{N}$, we have

$$\lambda_{\min}\left(n^{-1}\mathrm{Var}_{\theta^*}^{\pi}(S_{n,\mathrm{QML}}(\theta^*))\right) \geq c_0, \qquad \lambda_{\min}\left(n^{-1}\mathrm{Var}_{\theta^*}^{\pi}(S_{n,\mathrm{ML}}(\theta^*))\right) \geq c_0$$

and

$$\lambda_{\min}\left(F_{n,\mathrm{QML}}\right) \geq c_0, \qquad \lambda_{\min}\left(F_{n,\mathrm{ML}}\right) \geq c_0.$$

To (P1'): The condition is satisfied by the additional model assumptions.
To (CLT1) and (CLT2): Condition (2.16) is satisfied by assumption. As in the Poisson model, we cannot verify this condition analytically, but simulations reveal that (2.16) holds, if

$$\beta_n = o(n^{-1/2}).$$

We refer to Section 6 for more details.

For simplicity, we will assume in the following that $M = 1$. The case $M > 1$ can be shown similarly by replacing the one-dimensional Gaussian density function with the M-dimensional Gaussian density function. For $j \in \{1, \ldots, d\}$, we use the notation $\Sigma_{\theta}^{(j)} = \sigma_{\theta}^{(j)}$. Recall that

$$\psi(\theta, s_1, s_2, z) = \frac{\partial}{\partial \theta} \log(P_{\theta}(s_1, s_2) f_{\theta}(s_2, z))$$

Fix an integer $r \in \{1, \ldots, d\}$. Assume that $\theta^{(r)} = P_{\theta}(j, k)$ for some $j, k \in S$. Then

$$\frac{\partial}{\partial \theta^{(r)}} \log(P_{\theta}(s_1, s_2) f_{\theta}(s_2, z)) = \frac{\mathbb{1}_{\{j\}}(s_1)\mathbb{1}_{\{k\}}(s_2)}{P_{\theta}(j, k)}.$$

Clearly, for such an r, we have that the sequence $(\psi^{(r)})_{i \in \mathbb{N}}$ belongs to the class C_3 by assumption

(P1'). Now, assume that $\theta^{(r)} = \mu_\theta^{(j)}$ for some $j \in S$, then

$$\frac{\partial}{\partial\theta^{(r)}} \log(P_\theta(s_1, s_2)f_\theta(s_2, z)) = \frac{(z - \mu_\theta^{(j)})\mathbb{1}_{\{j\}}(s_2)}{\left(\sigma_\theta^2\right)^{(j)}}.$$

Since for any $i \in \mathbb{N}$, Z_i is mixture of normally distributed random variables, it follows that $(\psi^{(r)})_{i\in\mathbb{N}}$ belongs to the class C_3. Assume now that $\theta^{(r)} = \left(\sigma_\theta^2\right)^{(j)}$ for some $j \in S$. we have that

$$\frac{\partial}{\partial\theta^{(r)}} \log(P_\theta(s_1, s_2)f_\theta(s_2, z)) = \frac{\mathbb{1}_{\{j\}}(s_2)}{2}\left(\frac{\left(z - \mu_\theta^{(j)}\right)^2}{\left(\sigma_\theta^4\right)^{(j)}} - \frac{1}{\left(\sigma_\theta^2\right)^{(j)}}\right).$$

Again, since for any $i \in \mathbb{N}$, Z_i is mixture of normally distributed random variables, it follows that $(\psi^{(r)})_{i\in\mathbb{N}}$ belongs to the class C_3. Assumption (CLT2) can be verified by similar arguments.

To (UC1) and (UC2): Note that the first condition of assumption (UC1) is satisfied by assumption. For $j \in \{1, \ldots, d\}$, we use again the notation $\Sigma_\theta^{(j)} = \sigma_\theta^{(j)}$. Similarly as above one can show that for $r, s = 1, \ldots, d$ we have $(\psi^{(r)})_{i\in\mathbb{N}}$ belongs to the class C_5 and that $(\partial\psi^{(r)}/\partial\theta^{(s)})_{i\in\mathbb{N}}$ belongs to the class C_3.

Now, we fix two integers $r, s \in \{1, \ldots, d\}$. Assume that $\theta^{(r)} = P_\theta(j, k)$ for some $j, k \in S$. For $s \neq r$, we have $\partial\psi_i^{(r)}/\partial\theta^{(s)} = 0$. For $s = r$, it follows that

$$\frac{\partial}{\partial\theta^{(s)}}\psi^{(r)}(\theta, s_1, s_2, z) = \frac{-\mathbb{1}_{\{j\}}(s_1)\mathbb{1}_{\{k\}}(s_2)}{(P_\theta(j, k))^2}.$$

It follows that $(\partial\psi^{(r)}/\partial\theta^{(s)})_{i\in\mathbb{N}}$ belongs to the class $C_{3,2}$ with

$$\bar{\psi}(z) = \sup_{\theta\in B(\theta^*, \delta_0)} \frac{2}{(P_\theta(s_1, s_2))^3}.$$

Assume now that $\theta^{(r)} = \mu_\theta^{(j)}$ for some $j \in S$. Then $\partial\psi^{(r)}/\partial\theta^{(s)} = 0$ whenever $s \neq r$ or $\theta^{(s)} = \sigma_\theta^{(j)}$. For $s = r$ we have

$$\frac{\partial}{\partial\theta^{(s)}}\psi^{(r)}(\theta, s_1, s_2, z) = \frac{-\mathbb{1}_{\{j\}}(s_2)}{\left(\sigma_\theta^2\right)^{(j)}}.$$

It follows that $(\partial\psi^{(r)}/\partial\theta^{(s)})_{i\in\mathbb{N}}$ belongs to the class $C_{3,2}$ with

$$\bar{\psi}(z) = \sup_{\theta\in B(\theta^*, \delta_0)} \frac{1}{\left(\sigma_\theta^4\right)^{(j)}}.$$

The other cases can be treated similarly. Assumption (UC2) follows by similar arguments.

The application of Theorem 2.12 and Proposition 2.11 leads to the following result.

Corollary 3.4. *For any initial distribution $\nu \in \mathcal{P}(S)$, we have in the setting of the linear Gaussian*

model that

$$\sqrt{n}G_{n,\text{QML}}^{-1/2}F_{n,\text{QML}}(\theta_{v,n}^{\text{QML}} - \theta^*) \xrightarrow{\mathcal{D}} Z,$$

and

$$\lim_{n\to\infty}\sqrt{n}G_{n,\text{ML}}^{-1/2}F_{n,\text{ML}}(\theta_{v,n}^{\text{ML}} - \theta^*) \xrightarrow{\mathcal{D}} Z$$

as $n \to \infty$, where $Z \sim \mathcal{N}(0, I)$, $G_{n,\text{QML}}^{1/2}G_{n,\text{QML}}^{1/2} = G_{n,\text{QML}}$ and $G_{n,\text{ML}}^{1/2}G_{n,\text{ML}}^{1/2} = G_{n,\text{ML}}$.

3.3 Discussion

Here we want to illustrate a hybrid model, i.e., the non-observed sequence Y is Poisson distributed and the inhomogeneous noise is normally distributed.

More precise, let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with irreducible transition matrix P_{θ^*} and stationary measure π. Assume that $X_1 \sim \pi$ and for $i = 1, \dots, K$ let $\lambda_{\theta^*}^{(i)} > 0$. Further, define the vector $\lambda_{\theta^*} = (\lambda_{\theta^*}^{(1)}, \dots, \lambda_{\theta^*}^{(K)})$. Conditioned on X the non-observed homogeneous sequence $Y = (Y_n)_{n\in\mathbb{N}}$ is an independent sequence of Poisson-distributed random variables with parameter $\lambda_{\theta^*}^{(X_n)}$. Hence, given X_n we have $Y_n \sim \text{Poi}(\lambda_{\theta^*}^{(X_n)})$. The observed sequence $Z = (Z_n)_{n\in\mathbb{N}}$ is determined by

$$Z_n = Y_n + \varepsilon_n,$$

where $(\varepsilon_n)_{n\in\mathbb{N}}$ is an independent sequence of random variables with $\varepsilon_n \sim \mathcal{N}(0, \beta_n^2)$ and a sequence $(\beta_n)_{n\in\mathbb{N}}$ is a sequence of positive and real-valued numbers, which converges sufficiently fast to zero.

The main issue here is that the observed sequence Z takes values in $\mathbb{R}$ whereas Y takes values in $\mathbb{N}$. Set $G = \mathbb{R}$ equipped with the reference measure

$$\lambda(A) = \mathcal{L}(A) + \sum_{i=0}^{\infty} \delta_i(A), \quad A \in \mathscr{B}(\mathbb{R}).$$

Here $\mathcal{L}(\cdot)$ denotes the Lebesgue measure and $\delta_i(\cdot)$ the Dirac-measure at point $i \in \mathbb{N}$. The conditional density $f_{\theta,n}(s, z)$ with respect to λ is given by

$$f_{\theta,n}(s,z) = \begin{cases} \sum_{j=0}^{\infty} \frac{\lambda_{\theta^*}^{(s)}}{j!}\exp(-\lambda_{\theta^*}^{(s)})\frac{1}{(2\pi\beta_n^2)^{1/2}}\exp\left(-\frac{(z-j)^2}{2\beta_n^2}\right) & z \in \mathbb{R}\setminus\mathbb{N} \\ 0 & z \in \mathbb{N}. \end{cases}$$

It is straightforward to show that (C2) is not satisfied in this scenario. Assumption (C2) is difficult to handle, whenever the support of f_θ is strictly "smaller" than the support of $f_{\theta,n}$.

We just want to mention a possible strategy to resolve this problem. First, transform the observed sequence Z to a sequence $\tilde{Z}$ such that the support of the conditional density $\tilde{f}_{\theta,n}$ is the same as the support of f_θ. In the illustrating Poisson model with Gaussian noise one can project the sequence to the natural numbers. Next, prove for this new model that the quasi-likelihood estimator $\tilde{\theta}_{v,n}^{\text{QML}}$ for $\tilde{Z}$ is consistent, for example by verifying the structural conditions above.

Finally prove that

$$\theta_{\nu,n}^{\mathrm{QML}} - \tilde{\theta}_{\nu,n}^{\mathrm{QML}} \to 0 \quad \mathbb{P}_{\theta^*}^{\pi}\mathrm{a.s.}$$

as $n \to \infty$. A similar strategy can be used to prove strong consistency for the maximum likelihood estimator.

Proofs of asymptotic results

In this section we will provide the strategy of the proofs of our main results. We relate the strategies to other proofs of asymptotic results for maximum likelihood estimation in HMMs and sketch the main steps. Details of technical proofs can be found in Appendix A.

4.1   Proof of Theorem 2.6

The general strategy of the proof is similar to the study of consistency of the MLE in homogeneous HMMs, see Baum and Petrie (1966), Leroux (1992) and Douc et al. (2011). It is based on the ideas in Wald (1949), i.e., we want to prove that for any closed set $C \subset \Theta$ with $\theta^* \notin C$

$$\mathbb{P}_{\theta^*}^{\pi} \left(\lim_{n \to \infty} \frac{\sup_{\theta \in C} q_{\theta}^{\nu}(Z_1, \dots, Z_n)}{q_{\theta^*}^{\nu}(Z_1, \dots, Z_n)} = 0 \right) = 1. \tag{4.1}$$

Recall that

$$\theta_{\nu,n}^{\mathrm{QML}} = \operatorname*{argmax}_{\theta \in \Theta} \log \left(q_{\theta}^{\nu}(z_1, \dots, z_n) \right).$$

It follows that

$$\frac{q_{\theta_{\nu,n}^{\mathrm{QML}}}^{\nu}(z_1, \dots, z_n)}{q_{\theta^*}^{\nu}(z_1, \dots, z_n)} \geq 1 \quad \forall n \in \mathbb{N}. \tag{4.2}$$

Given (4.1) and (4.2), Theorem B.1 shows the strong consistency of θ_n^{QML}. In order to show (4.1), Lemma B.2 implies that it is sufficient to prove that

$$\limsup_{n \to \infty} \sup_{\theta \in C} \frac{1}{n} \ell_{\nu,n}^{\mathrm{Q}}(\theta) < \lim_{n \to \infty} \frac{1}{n} \ell_{\nu,n}^{\mathrm{Q}}(\theta^*), \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.}, \tag{4.3}$$

provided the limit on the right side exists, which will be shown in Theorem 4.5. The basic idea to show (4.3) is to prove that the process Z is asymptotically mean stationary (a.m.s.) with stationary mean $\mathbb{P}_{\theta^*}^{\pi,Y}$. We refer to Definition 4.2 for a precise definition. The a.m.s. property enables us to use ergodic theory for the process Z. This in combination with results in the homogeneous case are the key tools. In Douc et al. (2011) the consistency of the MLE in homogeneous HMMs is verified under weak conditions. We use the following result of them, which verifies that the

relative entropy rate exists.

Theorem 4.1. *(Douc et al., 2011, Theorem 9) Assume that conditions (P1) and (H1) are satisfied. Then, there exists an $\ell(\theta^*) \in \mathbb{R}$, such that*

$$\ell(\theta^*) = \lim_{n \to \infty} \mathbb{E}_{\theta^*}^{\pi} \left[n^{-1} \log q_{\theta^*}^{\pi}(Y_1, \ldots, Y_n) \right] \tag{4.4}$$

and

$$\ell(\theta^*) = \lim_{n \to \infty} n^{-1} \log q_{\theta^*}^{\nu}(Y_1, \ldots, Y_n), \quad \mathbb{P}_{\theta^*}^{\pi}\text{-}a.s. \tag{4.5}$$

for any probability measure $\nu \in \mathcal{P}(S)$ which is strictly positive if and only if π is strictly positive.

In the proof of the previous result one essentially uses the generalized Shannon-McMillan-Breiman theorem for stationary processes proven by Barron (1985). Additionally, we also use a version of the generalized Shannon-McMillan-Breiman theorem for asymptotic mean stationary processes, also proven in Barron (1985). In the following we provide basic definitions to apply this result, for a detailed survey let us refer to Gray (2009).

Definition 4.2. Let $(\Omega, \mathscr{F})$ be a measurable space equipped with a probability measure $\mathbb{Q}$ and let $T : \Omega \to \Omega$ be a measurable mapping. Then

- $\mathbb{Q}$ is *ergodic*, if for every $A \in \mathcal{I}$ either $\mathbb{Q}(A) = 0$ or $\mathbb{Q}(A) = 1$. Here $\mathcal{I}$ denotes the σ-algebra of the invariant sets, that are, the sets $A \in \mathscr{F}$ satisfying $T^{-1}(A) = A$.

- $\mathbb{Q}$ is called *asymptotically mean stationary* (a.m.s.) if there is a probability measure $\bar{\mathbb{Q}}$ on $(\Omega, \mathscr{F})$, such that for all $A \in \mathscr{F}$ we have

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{Q}\left(T^{-j}A\right) \to \bar{\mathbb{Q}}(A),$$

  as $n \to \infty$. We call $\bar{\mathbb{Q}}$ *stationary mean* of $\mathbb{Q}$.

- a probability measure $\widehat{\mathbb{Q}}$ on $(\Omega, \mathscr{F})$ *asymptotically dominates* $\mathbb{Q}$ if for all $A \in \mathscr{F}$ with $\widehat{\mathbb{Q}}(A) = 0$ holds

$$\lim_{n \to \infty} \mathbb{Q}\left(T^{-n}A\right) = 0.$$

We need the following equivalence from Rechard (1956). The result follows also by virtue of Theorem 2, Theorem 3 and the remark after Theorem 3 in Gray and Kieffer (1980).

Lemma 4.3. *Let $(\Omega, \mathscr{F}, \mathbb{Q})$ be a probability space and $T : \Omega \to \Omega$ be a measurable mapping. Then, the following statements are equivalent:*

(i) The probability measure $\mathbb{Q}$ is a.m.s. with stationary mean $\bar{\mathbb{Q}}$.

(ii) There is a stationary probability measure $\widehat{\mathbb{Q}}$, which asymptotically dominates $\mathbb{Q}$.

In our inhomogeneous HMM situation $(\Omega, \mathscr{F})$ is the space $G^{\mathbb{N}}$ generated by the one-sided sequence $Z = (Z_1, Z_2, \ldots)$ equipped with the product σ-field $\mathscr{B} = \bigotimes_{i \in \mathbb{N}} \mathscr{B}(G)$. The transformation $T \colon G^{\mathbb{N}} \to G^{\mathbb{N}}$ is the left time shift, that is, for $A \in \mathscr{B}$ and $i \in \mathbb{N}$ we have

$$T^{-i}(A) = \left\{ (z_1, z_2, \ldots) \in G^{\mathbb{N}} : (z_{1+i}, z_{2+i}, \ldots) \in A \right\}. \tag{4.6}$$

Finally $\mathbb{Q} = \mathbb{P}_{\theta^*}^{\pi, Z}$. In this setting we have the following result:

Theorem 4.4. *Let us assume that condition (C1) is satisfied. Then* $\mathbb{P}_{\theta^*}^{\pi, Z}$ *is a.m.s. with stationary mean* $\mathbb{P}_{\theta^*}^{\pi, Y}$.

Proof. See Appendix A. □

Theorem 4.5. *Assume that the conditions (P1), (H1), (H4), (C1) and (C2) are satisfied. Then*

$$\lim_{n \to \infty} n^{-1} \log q_{\theta^*}^{\nu}(Z_1, \ldots, Z_n) = \ell(\theta^*) \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.}$$

for any probability measure $\nu \in \mathcal{P}(S)$ *which is strictly positive if and only if* π *is strictly positive.*

Proof. See Appendix A. □

While most of the previous work consider the relative entropy $\ell(\theta)$ (here $\ell(\theta)$ is defined analogously to $\ell(\theta^*)$), for each $\theta \in \Theta$ and prove that the relative distance $\ell(\theta^*) - \ell(\theta)$ is bounded away from 0, Douc et al. (2011) considered a more direct approach which does not involve the convergence of the relative entropy for each $\theta \in \Theta$. Now, we provide a lemma which is essentially used and proven in Douc et al. (2011). In our setting the formulation and the statement slightly simplifies compared their result, since we only consider finite state spaces.

Lemma 4.6. *Let δ be the counting measure on S. Assume that the conditions (P1), (P2) and (H1) – (H3) are satisfied. Then, for any $\theta \in \Theta$ with $\theta \nsim \theta^*$, there exists a natural number n_θ and a real number $\eta_\theta > 0$ such that $B(\theta, \eta_\theta) \subseteq \mathcal{U}_\theta$ and*

$$\frac{1}{n_\theta} \mathbb{E}_{\theta^*}^{\pi} \left[\sup_{\theta' \in B(\theta, \eta_\theta)} \log q_{\theta'}^{\delta}(Y_1, \ldots, Y_{n_\theta}) \right] < \ell(\theta^*). \tag{4.7}$$

Here $B(\theta, \eta) \subseteq \Theta$ is the Euclidean ball of radius $\eta > 0$ centered at $\theta \in \Theta$.

Proof. The result follows straightforward from Theorem 12 and the arguments in the proof of Lemma 13 in Douc et al. (2011). □

With Theorem 4.4, Theorem 4.5 and Lemma 4.6, we can finally show the strong consistency result.

Theorem 4.7. *Assume that the irreducibility and continuity conditions (P1), (P2), the closeness conditions (C1), (C2) and the well behaving HMM conditions (H1) – (H4) are satisfied. Further, let the initial distribution $\nu \in \mathcal{P}(S)$ be strictly positive if and only if π is strictly positive. Then*

$$\limsup_{n \to \infty} \sup_{\theta \in C} \frac{1}{n} \ell_{\nu, n}^{\mathbb{Q}}(\theta) < \ell(\theta^*), \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.},$$

Proof. See Appendix A.   □

4.2   Proof of Corollary 2.7

In this section we will show that under the assumptions of Theorem 2.6 and condition (C3) the relative entropy for each $\theta \in \Theta$ with respect to p_θ^ν is "close" to the relative entropy with respect to q_θ^ν for any $\nu \in \mathcal{P}(S)$. Using the same strategy from the previous section this implies the strong consistency of $\theta_{n,\nu}^{\mathrm{ML}}$ whenever $\theta_{n,\nu}^{\mathrm{QML}}$ is strongly consistent.

Proof of Corollary 2.7.   We use the same strategy as in the proof of Theorem 2.6. By Theorem A.7 it follows that

$$\lim_{n\to\infty} n^{-1} \log p_{\theta^*}^\pi(Z_1, \ldots, Z_n) = \ell(\theta^*) \quad \mathbb{P}_{\theta^*}^\pi\text{-a.s.}$$

For $\theta \nsim \theta^*$, we chose $\kappa_\theta \leq \eta_\theta$, where η_θ is defined in Lemma 4.6, such that $B(\theta, \kappa_\theta) \subset \mathcal{E}_\theta$. As explained in the proof of Theorem 2.6, it is sufficient to verify that for any closed set $C \subseteq \Theta$ with $\theta^* \notin C$ we have

$$\limsup_{n\to\infty} \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} n^{-1} \log p_{\theta'}^\nu(Z_1, \ldots, Z_n) < \ell(\theta^*) \quad \mathbb{P}_\theta^\pi\text{-a.s.} \tag{4.8}$$

With $k \in \mathbb{N}$ from condition (C3) we obtain by using (2.13) that

$$\limsup_{n\to\infty} \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} n^{-1} \log \left(\frac{p_{\theta'}^\nu(Z_1, \ldots, Z_n)}{q_{\theta'}^\nu(Z_1, \ldots, Z_n)} \right)$$

$$\leq \limsup_{n\to\infty} \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} n^{-1} \log \left(\prod_{i=1}^n \max_{s\in S} \frac{f_{\theta',i}(s, Z_i)}{f_{\theta'}(s, Z_i)} \right)$$

$$= \limsup_{n\to\infty} \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} n^{-1} \log \left(\prod_{i=k}^n \max_{s\in S} \frac{f_{\theta',i}(s, Z_i)}{f_{\theta'}(s, Z_i)} \right)$$

$$\leq \limsup_{n\to\infty} n^{-1} \log \left(\prod_{i=k}^n \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} \max_{s\in S} \frac{f_{\theta',i}(s, Z_i)}{f_{\theta'}(s, Z_i)} \right).$$

By the same arguments as for proving (A.8) in the proof of Theorem 4.5 we get that

$$\mathbb{P}_{\theta^*}^\pi \left(n^{-1} \log \left(\prod_{i=k}^n \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} \max_{s\in S} \frac{f_{\theta',i}(s, Z_i)}{f_{\theta'}(s, Z_i)} \right) \geq \varepsilon \right) \leq \exp\left(n(c_n - \varepsilon)\right),$$

with

$$c_n := \limsup_{n\to\infty} n^{-1} \log \left(\mathbb{E}_{\theta^*}^\pi \left[\prod_{i=k}^n \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} \max_{s\in S} \frac{f_{\theta',i}(s, Z_i)}{f_{\theta'}(s, Z_i)} \right] \right).$$

Corollary A.9 and the Borel Cantelli lemma imply that

$$\mathbb{P}_{\theta^*}^\pi \left(\limsup_{n\to\infty} \sup_{\theta' \in B(\theta,\kappa_\theta)\cap C} n^{-1} \log \left(\frac{p_{\theta'}^\nu(Z_1, \ldots, Z_n)}{q_{\theta'}^\nu(Z_1, \ldots, Z_n)} \right) \leq 0 \right) = 1.$$

Similarly, it follows that

$$\mathbb{P}^\pi_{\theta^*}\left(\limsup_{n\to\infty}\ \sup_{\theta'\in B(\theta,\kappa_\theta)\cap C}\ n^{-1}\log\left(\frac{q^\gamma_{\theta'}(Z_1,\ldots,Z_n)}{p^\gamma_{\theta'}(Z_1,\ldots,Z_n)}\right)\le 0\right)=1,$$

which implies that

$$\limsup_{n\to\infty}\ \sup_{\theta'\in B(\theta,\kappa_\theta)\cap C}\ n^{-1}\log p^\gamma_{\theta'}(Z_1,\ldots,Z_n)=\limsup_{n\to\infty}\ \sup_{\theta'\in B(\theta,\kappa_\theta)\cap C}\ n^{-1}\log q^\gamma_{\theta'}(Z_1,\ldots,Z_n).$$

Finally, the assertion follows from (A.11). $\square$

4.3 Proof of Theorem 2.12

In this section we will sketch the main steps of the proof of Theorem 2.12. The proof is closely related to the proof of Theorem 1 in Jensen (2011a) and consists of two steps. First, in Theorem 4.8 a central limit theorem for the $S_{n,\mathrm{QML}}(\theta^*)$ is proven. Second, Theorem 4.9 shows that the derivative of $S_{n,\mathrm{QML}}(\theta^*)$ converges to a non-random limit. A Taylor expansion then yields the asymptotic normality of any strongly consistent estimator. This strategy is widely used in proving asymptotic normality of the MLE in homogeneous HMMs, see Bickel et al. (1998), Douc and Matias (2001) and Douc et al. (2004).

Theorem 4.8. *Suppose that (P1') and (CLT1) hold. Then*

$$\mathrm{Var}^\pi_{\theta^*}\left(S_{n,\mathrm{QML}}(\theta^*)\right)^{-1/2}S_{n,\mathrm{QML}}(\theta^*)\overset{D}{\to} Z,$$

as $n\to\infty$, where $Z\sim\mathcal{N}(0,I)$.

Theorem 4.9. *Recall the definition of $F_{n,\mathrm{QML}}$ from (2.17). Suppose that (P1') and (UC1) hold. Let $J_n(\theta)=-\nabla S_{n,\mathrm{QML}}(\theta)$. Then*

$$\lim_{n\to\infty}\ \sup_{\theta\in B(\theta^*,\delta_n)}\left|J_n(\theta)/n-F_{n,\mathrm{QML}}\right|\overset{\mathbb{P}^\pi_{\theta^*}}{\longrightarrow}0,$$

as $n\to\infty$ for any real-valued sequence $(\delta_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\delta_n=0$.

Given Theorem 4.8 and Theorem 4.9 the proof of Theorem 2.12 is straightforward. Let $a,b,n\in\mathbb{N}$ with $a\le b$. For legibility reasons we occasionally write $w_{a:b}$ instead of $w_a,\ldots,w_b$ for arbitrary sequences. In the following let ∇ and ∇^2 take derivatives w.r.t. $\theta\in\Theta$. Equation (2.9)

implies that for some sequence $(\bar{\theta}_n)_{n\in\mathbb{N}}$ in Θ with $\bar{\theta}_n \to \theta^*$ $\mathbb{P}_{\theta^*}^{\pi}$-a.s. as $n \to \infty$ we have

$$
\begin{aligned}
0 &= \nabla \log q^{\nu}_{\theta^{\mathrm{QML}}_{\nu,n}}(Z_1,\ldots,Z_n) \\
&= \nabla \log q^{\nu}_{\theta^*}(Z_1,\ldots,Z_n) + \nabla^2 \log q^{\nu}_{\bar{\theta}_n}(Z_1,\ldots,Z_n)(\theta^{\mathrm{QML}}_{\nu,n} - \theta^*) \\
&= S_{n,\mathrm{QML}}(\theta^*) - J_n(\bar{\theta}_n)(\theta^{\mathrm{QML}}_{\nu,n} - \theta^*) \\
&\quad + \mathbb{E}^{\nu}_{\theta^*}\left[\nabla \log\left(\nu(X_1)f_{\theta^*}(X_1,Z_1)\right)\frac{q^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})} \mid Z_{1:n}\right] \\
&\quad + \nabla\mathbb{E}^{\nu}_{\theta^*}\left[\nabla \log\left(\nu(X_1)f_{\theta^*}(X_1,Z_1)\right)\frac{q^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,1}(X_1 \mid Z_{1:n})} \mid Z_{1:n}\right](\theta^{\mathrm{QML}}_{\nu,n} - \theta^*).
\end{aligned}
$$

Note that

$$
\left\|\left(\mathbb{E}^{\nu}_{\theta^*}\left[\nabla \log\left(\nu(X_1)f_{\theta^*}(X_1,Z_1)\right)\frac{q^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})} \mid Z_{1:n}\right]\right)\right\| \leq \max_{s\in S}\left|\nabla \log f_{\theta^*}(s,Z_1)\right| < \infty,
$$

and

$$
\left\|\nabla\left(\mathbb{E}^{\nu}_{\theta^*}\left[\nabla \log\left(\nu(X_1)f_{\theta^*}(X_1,Z_1)\right)\frac{q^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})} \mid Z_{1:n}\right]\right)\right\|_2 \leq \max_{s\in S}\left|\nabla^2 \log f_{\theta^*}(s,Z_1)\right| < \infty,
$$

by assumption (CLT1). Furtermore, assumption (CLT1) implies that

$$
\lambda_{\min}\left(\mathrm{Var}^{\pi}_{\theta^*}(S_{n,\mathrm{QML}})(\theta^*)\right) \geq nc_0
$$

for some constant $c_0 > 0$. It follows that

$$
\mathrm{Var}^{\pi}_{\theta^*}(S_{n,\mathrm{QML}}(\theta^*))^{-1/2}\left(\mathbb{E}^{\nu}_{\theta^*}\left[\nabla \log\left(\nu(X_1)f_{\theta^*}(X_1,Z_1)\right)\frac{q^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})} \mid Z_{1:n}\right]\right) \xrightarrow{\mathbb{P}^{\pi}_{\theta^*}} 0
$$

and

$$
\mathrm{Var}^{\pi}_{\theta^*}(S_{n,\mathrm{QML}}(\theta^*))^{-1/2}\nabla\left(\mathbb{E}^{\nu}_{\theta^*}\left[\nabla \log\left(\nu(X_1)f_{\theta^*}(X_1,Z_1)\right)\frac{q^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,1:1}(X_1 \mid Z_{1:n})} \mid Z_{1:n}\right]\right) \xrightarrow{\mathbb{P}^{\pi}_{\theta^*}} 0,
$$

as $n \to \infty$. This together with Theorem 4.8 imply that

$$
\mathrm{Var}^{\pi}_{\theta^*}(S_{n,\mathrm{QML}}(\theta^*))^{-1/2} J_n(\bar{\theta}_n)(\theta^{\mathrm{QML}}_{\nu,n} - \theta^*) \to Z,
$$

where $Z \sim \mathcal{N}(0, I)$ as $n \to \infty$. Finally, note that

$$\sqrt{n} G_{n,QML}^{-1/2} F_{n,QML}(\theta_{v,n}^{\text{QML}} - \theta^*)$$

$$= n \text{Var}_{\theta^*}^{\pi} (S_{n,QML}(\theta^*))^{-1/2} J_n(\bar{\theta}_n)/n \left(J_n(\bar{\theta}_n)/n \right)^{-1} F_{n,QML}(\theta_{v,n}^{\text{QML}} - \theta^*)$$

$$= \text{Var}_{\theta^*}^{\pi} (S_{n,QML}(\theta^*))^{-1/2} J_n(\bar{\theta}_n) \left(J_n(\bar{\theta}_n)/n \right)^{-1} F_{n,QML}(\theta_{v,n}^{\text{QML}} - \theta^*)$$

$$\xrightarrow{D} \mathcal{N}(0, I_d)$$

as $n \to \infty$ by Theorem 4.9 and Slutsky's theorem.

The main difficulty in proving Theorem 4.8 and Theorem 4.9 arises by replacing the conditional density $f_{\theta,i}$ with f_θ, since the expected value of $S_{n,\text{QML}}(\theta^*)$ is not zero.

4.3.1   A central limit theorem

In order to derive a central limit theorem (CLT) for $S_{n,\text{QML}}(\theta^*)$ we will use a CLT for sums of weakly dependent random variables developed by Jensen (2011b).

Theorem 4.10 (Theorem 1 in Jensen (2011b)). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(W_i)_{i \in \mathbb{N}}$ be random vectors in $\mathbb{R}^d, d \in \mathbb{N}$ and $S_n = \sum_{i=1}^n W_i$. Let $(\mathcal{D}_j)_{j \in \mathbb{N}}$ be a set of σ-algebras. For an index set $I_1 \subset \mathbb{N}$ let $\sigma(D_i, i \in I_1)$ be the smallest σ-algebra that contains all sets $A \in D_i, i \in I_1$. For two index sets I_1, I_2 with $I_1, I_2 \subset \mathbb{N}$, define the strong mixing coefficient by*

$$\alpha(k, l, r) = \sup_{\substack{A_i \in \sigma(D_j : j \in I_i), \ i=1,2 \\ |I_2| \leq k, \ |I_1| \leq l, \ \text{dist}(I_1, I_2) \geq r}} |\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)|.$$

Here

$$\text{dist}(I_1, I_2) = \min_{i \in I_1, j \in I_2} |i - j|.$$

Assume that there exist constants $\delta_0, \varepsilon_0 > 0$, constants $\delta_1, \delta_2 \geq 0$, a constant β with $\beta > \delta_1 + \delta_2 + \max\{(2 + \delta_0)/\delta_0, 1 + \delta_2, 2\}$ and constants c_0, c_1, c_2 such that the following holds:

(1) *For all $i \in \mathbb{N}$ we have $\mathbb{E}[W_i] = 0$ and $\mathbb{E}\left[|W_i|^{2+\delta_0}\right] \leq c_0$. Further, assume that there exists an integer n_0 such that for all $n \geq n_0, n \in \mathbb{N}$, we have*

$$\text{Var}(a^T S_{n,QML}) \geq \varepsilon_0 n |a| \quad \forall a \in \mathbb{R}^d,$$

(2) *For $k, l, r \in \mathbb{N}$ we have $\alpha(k, l, r) \leq c_1 k^{\delta_1} l^{\delta_2} \max\{1, r\}^{-\beta}$.*

(3) *For all $r \in \mathbb{N}$ there exists a random variable W_j^r which is measurable w.r.t. $\sigma(D_k : |k - j| \leq r)$ and $\mathbb{E}\left[\left|W_j - W_j^r\right|\right] \leq c_2 r^{-\beta}$.*

Then we have

$$\text{Var}(S_n)^{-1/2} S_n \xrightarrow{d} Z,$$

where $Z \sim \mathcal{N}(0, I)$ as $n \to \infty$.

In order to apply Theorem 4.10 we need a strong mixing result for the observation sequence Z.

Lemma 4.11 (Corollary 1 in Jensen (2011a)). *Assume condition (P1') holds with constants δ_0 and p_0 and set $\rho = 1 - p_0^2$. For any integers r, t with $r < t$, any $\theta \in B(\theta^*, \delta_0)$ and any $B \in \mathcal{S}$ we have*

$$\sup_{s' \in S} \mathbb{P}_\theta^\pi (X_t \in B \mid X_r = s', Z_1, \dots, Z_n) - \inf_{s' \in S} \mathbb{P}_\theta^\pi (X_t \in B \mid X_r = s', Z_1, \dots, Z_n) \le \rho^{t-r}.$$

Furthermore, for any integers r, l, t_1, t_2 with $r < t_1$ and $t_2 < l$, any $\theta \in B(\theta^*, \delta_0)$ and any $B \in \mathcal{S}^{t_2 - t_1 + 1}$ we have

$$\sup_{s_l, s_r \in S} \mathbb{P}_\theta^\pi ((X_{t_1}, \dots, X_{t_2}) \in B \mid X_r = s_r, X_l = s_l, Z_1, \dots, Z_n)$$
$$- \inf_{s_l, s_r \in S} \mathbb{P}_\theta^\pi ((X_{t_1}, \dots, X_{t_2}) \in B \mid X_r = s_r, X_l = s_l, Z_1, \dots, Z_n)$$
$$\le \rho^{t_1 - r} + \rho^{l - t_2}.$$

Recall the definition of $q_{\theta, I_2}^{\nu, I_1}$ in (2.5). Now, we define a similar function which includes conditioning on the underlying Markov chain as well. To this end let $n \in \mathbb{N}$ and I_1, I_2, I_3 be finite ordered sets with $I_2, I_3 \subset I_1 \subset \{1, \dots, n\}$. Further let $s \in S^{I_2}$, $u \in S^{I_3}$ and $z \in G^{I_1}$. We set

$$q_{\theta, I_2 | I_3}^{\nu, I_1} (s \mid u, z) = \frac{\displaystyle\int_{\substack{y = (y_1, \dots, y_n) \in G^n: \\ y_{|I_1} = z}} \sum_{\substack{x = (x_1, \dots, x_n) \in S^n: \\ x_{|I_2} = s, x_{|I_3} = u}} \nu(x_1) f_\theta(x_1, y_1) \prod_{i=2}^n P_\theta(x_{i-1}, x_i) f_\theta(x_i, y_i) \lambda_n(y)}{\displaystyle\int_{\substack{y = (y_1, \dots, y_n) \in G^n: \\ y_{|I_1} = z}} \sum_{\substack{x = (x_1, \dots, x_n) \in S^n: \\ x_{|I_3} = u}} \nu(x_1) f_\theta(x_1, y_1) \prod_{i=2}^n P_\theta(x_{i-1}, x_i) f_\theta(x_i, y_i) \lambda_n(y)}.$$

For $a, r, l, b, i, j \in \mathbb{N}$ with $a \le i \le j \le b$ and $a \le r \le l \le b$ we write $q_{\theta, i: j | r: l}^{\nu, a:b}$ for $q_{\theta, \{i, \dots, j\} | \{r, \dots, l\}}^{\nu, \{a, \dots, b\}}$. The following corollary can be shown similarly to Lemma 4.11 by replacing the conditional density $f_{\theta, i}$ with f_θ.

Corollary 4.12. *Assume condition (P1') holds with constants δ_0 and p_0 and set $\rho = 1 - p_0^2$. For any integers r, t with $r < t$, any $\theta \in B(\theta^*, \delta_0)$ and any $B \in \mathcal{S}$ we have*

$$\sup_{s' \in S} \sum_{x \in B} q_{\theta, t: t | r: r}^{\nu, 1:n} (x \mid s', Z_1, \dots, Z_n) - \inf_{s' \in S} \sum_{x \in B} q_{\theta, t: t | r: r}^{\nu, 1:n} (x \mid s', Z_1, \dots, Z_n) \le \rho^{t-r}.$$

Furthermore, for any integers r, l, t_1, t_2 with $r < t_1$ and $t_2 < l$, any $\theta \in B(\theta^*, \delta_0)$ and any

$B \in \mathcal{S}^{t_2-t_1+1}$ *we have*

$$\sup_{s_l,s_r \in S} \sum_{(x_{t_1},\ldots,x_{t_2}) \in B} q^{\nu,1:n}_{\theta,t_1:t_2|\{r,l\}}(x_{t_1},\ldots,x_{t_2} \mid s_r,s_l,Z_1,\ldots,Z_n)$$

$$- \inf_{s_l,s_r \in S} \sum_{(x_{t_1},\ldots,x_{t_2}) \in B} q^{\nu,1:n}_{\theta,t_1:t_2|\{r,l\}}(x_{t_1},\ldots,x_{t_2} \mid s_r,s_l,Z_1,\ldots,Z_n)$$

$$\leq \rho^{t_1-r} + \rho^{l-t_2}.$$

Lemma 4.13 (Corollary 2 in Jensen (2011a))**.** *Assume condition (P1') holds with constants δ_0 and p_0 and set $\rho = 1 - p_0^2$. For any integers r, s, t with $r < s < t$, any $\theta \in B(\theta^*, \delta_0)$ and any $B \in \mathcal{B}(G)$ we have*

$$\sup_{z_r,z_t \in G} \mathbb{P}^\pi_\theta(Z_s \in B \mid Z_r = z_r, Z_t = z_t) - \inf_{z_r,z_t \in G} \mathbb{P}^\pi_\theta(Z_s \in B \mid Z_r = z_r, Z_t = z_t) \leq \rho^{s-r} + \rho^{t-s}.$$

Corollary 4.14. *Assume condition (P1') holds with constants δ_0 and p_0 and set $\rho = 1 - p_0^2$. Let $\nu_1, \nu_2 \in \mathcal{P}(S)$ be two probability measures on S. For any integers n, r, l, i, with $n \geq l \geq i \geq r \geq 1$ with and any $\theta \in B(\theta^*, \delta_0)$ it holds that*

$$\sum_{s_{i-1},s_i \in S} \sum_{s_r,s_l \in S} \left| q^{\nu,r:l}_{\theta,(i-1):i|\{r,l\}}(s_{i-1},s_i \mid s_r,s_l,Z_r,\ldots Z_l) \right.$$

$$\times \left(q^{\nu,r:l}_{\theta,(i-1):i}(s_{i-1},s_i \mid Z_r,\ldots,Z_l) - q^{\nu,1:n}_{\theta,(i-1):i}(s_{i-1},s_i \mid Z_1,\ldots,Z_n) \right) \Big|$$

$$\leq 2(\rho^{i-r} + \rho^{l-i}).$$

Proof. We refer to Appendix A. $\square$

Proof of Theorem 4.8. Note that for any integer n with $n \geq n_0$ we have

$$S_{n,\text{QML}}(\theta^*) = \sum_{i=2}^n \mathbb{E}^\nu_{\theta^*} \left[\psi(\theta^*,X_{i-1},X_i,Z_i) \frac{q^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})} \mid Z_{1:n} \right]$$

$$= \sum_{i=2}^n \left(\mathbb{E}^\nu_{\theta^*} \left[\psi(\theta^*,X_{i-1},X_i,Z_i) \frac{q^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right.$$

$$\left. - \mathbb{E}^\pi_{\theta^*} \left[\mathbb{E}^\nu_{\theta^*} \left[\psi(\theta^*,X_{i-1},X_i,Z_i) \frac{q^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right] \right)$$

$$+ \sum_{i=2}^n \mathbb{E}^\pi_{\theta^*} \left[\mathbb{E}^\nu_{\theta^*} \left[\psi(\theta^*,X_{i-1},X_i,Z_i) \frac{q^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(i-1):i}(X_{i-1},X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right].$$

Observe that

$$
\left\| \mathrm{Var}_{\theta^*}^\pi (S_{n,\mathrm{QML}}(\theta^*))^{-1/2} \sum_{i=2}^n \mathbb{E}_{\theta^*}^\pi \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right] \right\|_1
$$

$$
= \left\| \mathrm{Var}_{\theta^*}^\pi \left(\frac{S_{n,\mathrm{QML}}(\theta^*)}{n} \right)^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=2}^n \mathbb{E}_{\theta^*}^\pi \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right] \right\|_1
$$

$$
\leq \left\| \mathrm{Var}_{\theta^*}^\pi \left(\frac{S_{n,\mathrm{QML}}(\theta^*)}{n} \right)^{-1/2} \right\|_1
$$

$$
\times \frac{1}{\sqrt{n}} \left\| \sum_{i=2}^n \mathbb{E}_{\theta^*}^\pi \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right] \right\|_1
$$

$$
\leq \sqrt{d} \left\| \mathrm{Var}_{\theta^*}^\pi \left(\frac{S_{n,\mathrm{QML}}(\theta^*)}{n} \right)^{-1/2} \right\|_2
$$

$$
\times \frac{1}{\sqrt{n}} \left| \sum_{i=2}^n \mathbb{E}_{\theta^*}^\pi \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right] \right\|_1
$$

$$
\leq \frac{\sqrt{d}}{\sqrt{c_0}} \frac{1}{\sqrt{n}} \left\| \sum_{i=2}^n \mathbb{E}_{\theta^*}^\pi \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right] \right\|_1,
$$

where we used Lemma B.25, the second part of assumption (CLT1) and the fact that for $A \in \mathbb{R}^{d \times d}$ it holds that

$$
\|A\|_1 \leq \sqrt{d} \, \|A\|_2 .
$$

Note that condition (2.16) in assumption (CLT1) implies that

$$
\mathrm{Var}_{\theta^*}^\pi (S_{n,QML}(\theta^*))^{-1/2} \sum_{i=1}^n \mathbb{E}_{\theta^*}^\pi \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right] \xrightarrow{D} 0,
$$

as $n \to \infty$.

Now, we apply Theorem 4.10. For $j, i \in \mathbb{N}$, let D_j be the σ-algebra generated by G. It follows that the strong mixing condition (2) is satisfied with $\delta_1 = \delta_2 = 0$ and replacing $\mathrm{dist}(I_1, I_2)^{-\beta}$ by $\rho^{\mathrm{dist}(I_1, I_2)}$ by Lemma 4.13 . We set

$$
W_i = \mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right]
$$

$$
- \mathbb{E}_{\theta^*}^\pi \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right].
$$

Let δ_0 and p_0 be the constants from assumption (P1') and set $\rho = 1 - p_0^2$. Further, condition

(CLT1) implies that for $r \in \{1, \dots, d\}$ we have that $(\psi^{(r)})_{i \in \mathbb{N}}$ belongs to the class C_3 and therefore

$$\mathbb{E}_\theta^\pi [W_i] = 0 \quad \text{and} \quad \mathbb{E}_\theta^\pi \left[|W_i|^3 \right] < \infty,$$

Now, we check the condition (3) of Theorem 4.10. Let $i, l \in \mathbb{N}$ such that $i - l \geq 1$ and $i + l \leq n$. Corollary 4.14 implies that

$$\left| \mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{(i-l):(i+l)} \right] \right.$$

$$\left. - \mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right|$$

$$= \left| \mathbb{E}_{\theta^*}^\nu \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,(i-l):(i+l)}(X_{i-1}, X_i \mid Z_{(i-l):(i+l)})}{p_{\theta^*,(i-1):i}^{\nu,(i-l):(i+l)}(X_{i-1}, X_i \mid Z_{(i-l):(i+l)})} \mid X_{i-l}, X_{i+l}, Z_{(i-l):(i+l)} \right] \mid Z_{(i-l):(i+l)} \right] \right.$$

$$\left. - \mathbb{E}_{\theta^*}^\nu \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid X_{i-l}, X_{i+l}, Z_{1:n} \right] \mid Z_{1:n} \right] \right|$$

$$= \left| \mathbb{E}_{\theta^*}^\nu \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,(i-l):(i+l)}(X_{i-1}, X_i \mid Z_{(i-l):(i+l)})}{p_{\theta^*,(i-1):i}^{\nu,(i-l):(i+l)}(X_{i-1}, X_i \mid Z_{(i-l):(i+l)})} \mid X_{i-l}, X_{i+l}, Z_{(i-l):(i+l)} \right] \mid Z_{(i-l):(i+l)} \right] \right.$$

$$\left. - \mathbb{E}_{\theta^*}^\nu \left[\mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{(i-l):(i+l)})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{(i-l):(i+l)})} \mid X_{i-l}, X_{i+l}, Z_{(i-l):(i+l)} \right] \mid Z_{1:n} \right] \right|$$

$$= \left| \sum_{s_{i-l}, s_{i+l}} \sum_{s_{i-1}, s_i} \psi(\theta^*, s_{i-1}, s_i, Z_i) q_{\theta,(i-1):i|\{i-l,i+l\}}^{\nu,(i-l):(i+l)}(s_{i-1}, s_i \mid s_{i-l}, s_{i+l}, Z_{i-l}, \dots, Z_{i+l}) \right.$$

$$\left. \times \left(q_{\theta,(i-1):i}^{\nu,(i-l):(i+l)}(s_{i-l}, s_{i+l} \mid Z_{i-l}, \dots, Z_{i+l}) - q_{\theta,(i-1):i}^{\nu,1:n}(s_{i-l}, s_{i+l} \mid Z_1, \dots, Z_n) \right) \right|$$

$$\leq \psi^0(Z_i) \sum_{s_{i-1}, s_i} \sum_{s_{i-l}, s_{i+l}} \left| q_{\theta,(i-1):i|\{i-l,i+l\}}^{\nu,(i-l):(i+l)}(s_{i-1}, s_i \mid s_{i-l}, s_{i+l}, Z_{i-l}, \dots, Z_{i+l}) \right.$$

$$\left. \times \left(q_{\theta,(i-1):i}^{\nu,(i-l):(i+l)}(s_{i-l}, s_{i+l} \mid Z_{i-l}, \dots, Z_{i+l}) - q_{\theta,(i-1):i}^{\nu,1:n}(s_{i-l}, s_{i+l} \mid Z_1, \dots, Z_n) \right) \right|$$

$$\leq \psi^0(Z_i) 2 \left(\rho^l + \rho^l \right) = 4 \psi^0(Z_i) \rho^l, \tag{4.9}$$

where $\psi^0 : G \to \mathbb{R}^d$ and $(\psi^0)^{(r)}$ is the bound of $\psi^{(r)}$ from Definition 2.3 for $r = 1, \dots, d$. Taking the expected value it follows from assumption (CLT1) that

$$\mathbb{E}_{\theta^*}^\pi \left[\left| \mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{(i-l):(i+l)} \right] \right. \right.$$

$$\left. \left. - \mathbb{E}_{\theta^*}^\nu \left[\psi(\theta^*, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \mid Z_{1:n} \right] \right| \right] \leq 4 \rho^l K^{1/3}$$

where K is an upper bound on the third moment of $(\psi^0)^{(r)}(Z_i)$ for all $i \in \mathbb{N}$ and all $r = 1, \dots, d$. The cases $i - l < 1$ and $i + l > n$ can be treated similarly using one-sided mixing.

$\square$

4.3.2   A uniform convergence of the observed information

In this section we will prove Theorem 4.9. The proofs are almost identical to the proofs of Section 5 in Jensen (2011a).

Recall that $J_n(\theta) = -\nabla S_{n,\mathrm{QML}}(\theta)$. Using formula (3.2) in Louis (1982) we find that

$$
\begin{aligned}
J_n(&\theta) \\
&= -\mathbb{E}_{\theta^*}^\nu \left[\sum_{i=2}^n \frac{\partial}{\partial \theta} \left(\psi(\theta, X_{i-1}, X_i, Z_i) \right) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \,\Big|\, Z_{1:n} \right] \\
&\quad - \mathrm{Var}_{\theta^*}^\nu \left(\sum_{i=2}^n \frac{\partial}{\partial \theta} \left(\psi(\theta, X_{i-1}, X_i, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{i-1}, X_i \mid Z_{1:n})} \,\Big|\, Z_{1:n} \right) \right).
\end{aligned}
$$

We will treat the expectation and the covariance matrix separately. For the next lemma, we define the function $h : G \to \mathbb{R}_+$ by

$$
h(z) = \sup_{\substack{s_i, s_2 \in S \\ \theta \in B(\theta^*, \delta_0), u \in \{1, \ldots, d\}}} \left| \psi^{(u)}(\theta, s_1, s_2, z) \right|. \tag{4.10}
$$

Recall that for $l, k \in \mathbb{N}$ with $l \geq k$, we use the abbreviation $w_{k:l} = w_k, \ldots, w_l$.

Lemma 4.15. *Assume condition (P1') holds with constants δ_0 and p_0 and let $b : S^{t-r+1} \times G^{t-r+1} \to \mathbb{R}$ be a function such that there is a function $b^0 : G^{t-r+1} \to \mathbb{R}_+$ with*

$$
\sup_{s_r, \ldots, s_t} b(s_{r:t}, z_{r:t}) \leq b^0(z_{r:t}).
$$

Then for any $\theta \in B(\theta^*, \delta_0)$ and any integer $k > 0$, we have

$$
\begin{aligned}
\left| \mathbb{E}_\theta^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})}{p_{\theta,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})} \,\Big|\, Z_{1:n} \right] - \mathbb{E}_{\theta^*}^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta^*,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})}{p_{\theta^*,r:t}^{\nu,1,n}(X_{r:t} \mid Z_{1:n})} \,\Big|\, Z_{1:n} \right] \right| \\
\leq 2 b^0(Z_{r:t}) \left(\rho \left| \theta - \theta^* \right| \sum_{i=r-k+1}^{t+k} h(Z_i) + 8 \rho^k \right)
\end{aligned}
$$

Proof. We refer to Appendix A.   □

Proposition 4.16 and Proposition 4.17 are essential in proving a uniform convergence result. Their proofs are based on Lemma 4.15.

Proposition 4.16. *Assume condition (P1') holds with constants δ_0 and p_0 and for $i \in \mathbb{N}$ let $a_i : \Theta \times S \times S \times G \to \mathbb{R}$ be a function and let $(a_i)_{i \in \mathbb{N}}$ belongs to $C_{2,1}$. Further let h be defined as in (4.10) and assume that there exists a constant K such that for all $i \in \mathbb{N}$ we have $\mathbb{E}_{\theta^*}^\pi \left[|h(Z_i)|^2 \right] < K$.*

Then, for any sequence $\delta_n \to 0$ as $n \to \infty$, we have

$$
\lim_{n\to\infty} \mathbb{E}_{\theta^*}^{\pi} \left[\sup_{\theta \in B(\theta^*, \delta_n)} \left| \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{E}_{\theta}^{\pi} \left[a_i(\theta, X_{(i-1):i}, Z_i) \frac{q_{\theta,(i-1):i}^{\gamma,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta,(i-1):i}^{\gamma,1:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n} \right] \right. \right. \right.
$$
$$
\left. \left. \left. - \mathbb{E}_{\theta^*}^{\pi} \left[a_i(\theta^*, X_{(i-1):i}, Z_i) \frac{q_{\theta^*,(i-1):i}^{\gamma,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\gamma,1:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n} \right] \right) \right| \right] = 0. \qquad (4.11)
$$

Proof. See Appendix A.   □

Proposition 4.17. *Assume condition (P1') holds with constants δ_0 and p_0 and for $i \in \mathbb{N}$ let $a_i, b_i : \Theta \times S \times S \times G \to \mathbb{R}$ be functions and let $(a_i)_{i \in \mathbb{N}}$ and $(b_i)_{i \in \mathbb{N}}$ belong to the class $C_{3,2}$. Further, let $(\delta_n)_{n \in \mathbb{N}}$ be a sequence such that $\delta_n \to 0$ as $n \to \infty$. Then*

$$
\mathbb{E}_{\theta^*}^{\pi} \left[\sup_{|\theta - \theta^*| \le \delta_n} \left| \frac{1}{n} \sum_{u,v=1}^{n} \left(\mathrm{Cov}_{\theta}^{\gamma} \left(a_u(\theta) \frac{q_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta) \frac{q_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta,(v-1):v}^{\gamma,1,n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right) \right. \right. \right.
$$
$$
\left. \left. \left. - \mathrm{Cov}_{\theta^*}^{\gamma} \left(a_u(\theta^*) \frac{q_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*) \frac{q_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right) \right) \right| \right] \to 0,
$$

as $n \to \infty$.

Proof. We refer to Appendix A.   □

Corollary 4.18. *Suppose that assumptions (P1') and (UC1) hold. Further, let $(\delta_n)_{n \in \mathbb{N}}$ be a sequence such that $\delta_n \to 0$ as $n \to \infty$. Then we have*

$$
\lim_{n\to\infty} \sup_{|\theta - \theta^*| > \delta_n} \left| \frac{1}{n} \left(J_n(\theta) - J_n(\theta^*) \right) \right| = 0.
$$

Now, we show that there exists a non-random matrix $F_{n,\mathrm{QML}} \in \mathbb{R}^{d \times d}$ such that the difference of $\frac{1}{n} J_n(\theta^*)$ and $F_{n,\mathrm{QML}}$ converges to zero.

Lemma 4.19. *Assume condition (P1') holds with constants δ_0 and p_0 and for $i \in \mathbb{N}$ let $a_i : \Theta \times S \times S \times G \to \mathbb{R}$ be a function and let $(a_i)_{i \in \mathbb{N}}$ belongs to C_3. Then*

$$
\lim_{n\to\infty} \mathrm{Var}_{\theta^*}^{\pi} \left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\theta^*}^{\pi} \left[a_i(\theta^*, X_{(i-1):i}, Z_i) \frac{q_{\theta^*,(i-1):i}^{\gamma,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\gamma,1:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n} \right] \right) = 0.
$$

Proof. We refer to Appendix A.   □

Lemma 4.20. *Assume condition (P1') holds with constants δ_0 and p_0 and for $i \in \mathbb{N}$ let $a_i, b_i : \Theta \times S \times S \times G \to \mathbb{R}$ be functions and let $(a_i)_{i \in \mathbb{N}}$ and $(b_i)_{i \in \mathbb{N}}$ belong to the class $C_{4+\delta}$, for some*

$\delta > 0$. *Then*

$$
\mathrm{Var}^\pi_{\theta^*}\left(\frac{1}{n}\sum_{u=1}^{n}\sum_{v=1}^{n}\mathrm{Cov}^\pi_{\theta^*}\left(a_u(\theta^*)\frac{q^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u}\mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u}\mid Z_{1:n})}, b_v(\theta^*)\frac{q^{\nu,1,n}_{\theta^*,(v-1):v}(X_{(v-1):v}\mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v}\mid Z_{1:n})}\mid Z_{1:n}\right)\right)
$$
$$
\to 0,
$$

as $n \to \infty$.

Proof. See Appendix A.   □

Corollary 4.21. *Suppose that assumptions (P1') and (UC1) hold. Then we have*

$$
\lim_{n\to\infty}\left|F_{n,\mathrm{QML}} - n^{-1}J_n(\theta^*)\right| = 0.
$$

4.4 Proof of Proposition 2.11

The assertion is a direct apllication of Theorem 1 in Jensen (2011a) and follows from assumptions (P1'), (CLT2) and (UC2).

Inference in hidden Markov models

In this section we provide an algorithm for calculating of the likelihood function and finding the MLE efficiently in HMMS. Given the setting described from Section 2.1 and observations $z_1, \ldots, z_n \in G$ for $n \in \mathbb{N}$ three basic problems are of interest:

(1) Given $\theta \in \Theta$ and an initial distribution ν for X_1, how can we compute the likelihood functions $p_\theta^\nu(z_1, \ldots, z_n)$ and $q_\theta^\nu(z_1, \ldots, z_n)$?

(2) For which $\theta \in \Theta$ is $\theta \mapsto p_\theta^\nu(z_1, \ldots, z_n)$ and $\theta \mapsto q_\theta^\nu(z_1, \ldots, z_n)$ maximal?

(3) Given $\theta \in \Theta$, what is the most likely corresponding sequence $s_1, \ldots s_n \in S$ of underlying states?

Remark 5.1. *Problem (1) – (3) were first handled by the work of Baum et al. (1970) and Viterbi (1967). Roughly described their idea is to use an iterative expectation-maximization (EM) algorithm which updates the parameter in each step and guarantees that the log-likelihood function ℓ_θ^ν is non-decreasing with respect to the updates of the parameter. After the iterative procedures is converged, the limiting value $\hat{\theta} \in \Theta$ can be used to determine the most likely underlying sequence for $\hat{\theta}$ beginning at X_n and going backwards. The methods of Baum and Viterbi were extended and specialized by various authors, e.g., Hsiao et al. (2009) and Gerber et al. (2011).*

Another approach solving the problems (2)-(3) are Markov chain Monte Carlo methods, especially Gibbs sampling. For a comparison of both methods we refer to Ryden (2008).

5.1 Computation of the likelihood function

In the following we will focus on the computation of $p_\theta^\nu(z_1, \ldots, z_n)$. The computation of $q_\theta^\nu(z_1, \ldots, z_n)$ can be treated similarly by replacing the time dependent conditional density $f_{\theta,i}$ with the time independent conditional density f_θ for $i = 1, \ldots, n$. Recall that the likelihood function can be computed via

$$p_\theta^\nu(z_1, \ldots, z_n) = \sum_{s_{1:n} \in S^n} \nu(s_1) f_{\theta,1}(s, z_1) \prod_{i=2}^n P_\theta(s_{i-1}, s_i) f_{\theta,i}(s, z_1). \tag{5.1}$$

We see that a direct computation of (5.1) has computational cost nK^n and is therefore not suitable for applications. The *forward-backward* algorithm, developed by (Baum and Eagon, 1967) has computational cost $O(nK^2)$ and is presented in the following.

The forward-backward algorithm

In this section we fix a parameter $\theta \in \Theta$ and an initial distribution $\nu \in \mathcal{P}(S)$ of the hidden Markov chain. First, we define the *forward* variables α_i and *backward* variables β_i, $i = 1, \ldots, n$. They can be computed recursively (see Proposition 5.3 below). We will suppress the dependency of α_i and β_i on θ and ν.

Definition 5.2. Let be $\alpha_1(s) = \pi(s) f_{\theta,1}(s, z_1)$ and $\beta_n(s) = 1 \ \forall s \in S$. Furthermore, for $i \in \{2, \ldots, n\}$ define

$$\alpha_i(s) := p_\theta^\nu(X_i = s, z_1, \ldots, z_i),$$

$$\beta_{i-1}(s) := p_\theta(z_i, \ldots, z_n \mid X_{i-1} = s),$$

where

$$p_\theta^\nu(X_i = s, z_1, \ldots, z_i) = \sum_{s_{1:(i-1)} \in S^{i-1}} \nu(s_1) f_{\theta,1}(s_1, z_1) \prod_{j=2}^{i-1} (P_\theta(s_{j-1}, s_j) f_{\theta,j}(s_j, z_j)) P_\theta(s_{i-1}, s) f_{\theta,i}(s, z_i)$$

and

$$p_\theta^\nu(z_i, \ldots, z_n \mid X_{i-1} = s) = \sum_{s_{i:n} \in S^{n-i+1}} P_\theta(s, s_i) f_{\theta,i}(s_i, z_i) \prod_{j=i+1}^{n} P_\theta(s_{j-1}, s_j) f_{\theta,j}(s_j, z_j).$$

Proposition 5.3. *For $i \in \{2, \ldots, n\}$ and $s \in S$ the forward and backward variables can be computed recursively via*

$$\alpha_i(s) = \sum_{j=1}^{K} \alpha_{i-1}(j) P_\theta(j, s) f_{\theta,i}(s, z_i), \tag{5.2}$$

and

$$\beta_{i-1}(s) = \sum_{j=1}^{K} \beta_i(j) P_\theta(s, j) f_{\theta,i}(j, z_i). \tag{5.3}$$

Proof. See Appendix A.   □

Now, we can rewrite the likelihood function.

Proposition 5.4. *Given $\theta \in \Theta$ and $\nu \in \mathcal{P}(S)$, we have*

$$p_\theta^\nu(z_1, \ldots, z_n) = \sum_{k=1}^{K} \sum_{j=1}^{K} \alpha_i(k) P_\theta(k, j) f_{\theta,i+1}(j, z_{i+1}) \beta_{i+1}(j), \quad i \in \{1, \ldots, n-1\}.$$

Proof. Let be $i \in \{1, \ldots, n-1\}$. Using the definitions of the *forward* and *backward* variables it follows that

$$p_\theta^\nu(z_1, \ldots, z_n)$$

$$= \sum_{s_{1:n} \in S^n} \nu(s_1) f_{\theta,1}(s_1, z_1) \prod_{t=2}^{n} P_\theta(s_{t-1}, s_t) f_{\theta,t}(s_t, z_t)$$

$$= \sum_{s_i \in S} \sum_{s_{i+1} \in S} \sum_{s_{1:i-1} \in S^{i-1}} \sum_{s_{i+2:n} \in S^{n-i-1}} \nu(s_1) f_{\theta,1}(s_1, z_1) \prod_{t=2}^{i-1} P_\theta(s_{t-1}, s_t) f_{\theta,t}(s_t, z_t) P_\theta(s_{i-1}, s_i) f_{\theta,i}(s_i, z_i)$$

$$P_\theta(s_i, s_{i+1}) f_{\theta,i+1}(s_{i+1}, z_{i+1}) \prod_{t=i+2}^{n} P_\theta(s_{t-1}, s_t) f_{\theta,t}(s_t, z_t)$$

$$= \sum_{k=1}^{K} \sum_{j=1}^{K} \sum_{s_{1:i-1} \in S^{i-1}} \nu(s_1) f_{\theta,1}(s_1, z_1) \prod_{t=2}^{i-1} P_\theta(s_{t-1}, s_t) f_{\theta,t}(s_t, z_t) P_\theta(s_{i-1}, k) f_{\theta,i}(k, z_i)$$

$$P_\theta(k, j) f_{\theta,i+1}(j, z_{i+1}) \sum_{s_{i+2:n} \in S^{n-i-1}} \prod_{t=i+2}^{n} P_\theta(s_{t-1}, s_t) f_{\theta,t}(s_t, z_t)$$

$$= \sum_{k=1}^{K} \sum_{j=1}^{K} \alpha_i(k) P_\theta(k, j) f_{\theta,i+1}(j, z_{i+1}) \beta_{i+1}(j)$$

$$\square$$

Remark 5.5. *The forward and backward variables can be computed with computational cost of* $O(nK^2)$.

5.2   Parameter estimation using dynamic programming

In this section we will approximate the MLE using an algorithm developed by Baum and Eagon (1967). In the HMM literature the algorithm is usually called *Baum-Welch* algorithm. It is an instance of the more general EM algorithm introduced by Dempster et al. (1977).

The expectation maximization algorithm

The EM algorithm is a general approach to the iterative computation of maximum likelihood estimates when the observations can be viewed as incomplete data. Hence, let $\mathscr{X}, \mathscr{Y}$ be two sample spaces and let $H : \mathscr{X} \to \mathscr{Y}$ be a surjective mapping. Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathscr{X}, \mathscr{B}(\mathscr{X}))$ and $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathscr{Y}, \mathscr{B}(\mathscr{Y}))$ be two random variables mapping from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into $\mathscr{X}, \mathscr{Y}$, respectively. The observed data $Y = y \in \mathscr{Y}$ corresponds to a least one $x \in \mathscr{X}$ via H, whereas $X = x$ is not observed. Additionally, assume that (Θ, m) is a Polish space and for $\theta \in \Theta$ the random variable X has a parametrized density function p_θ with respect to a

σ-finite λ measure on $\mathcal{X}$. Then, Y has a density function $g_\theta(\cdot)$, given by

$$g_\theta(y) = \int\limits_{\{x : H(x) = y\}} p_\theta(x)\lambda(\mathrm{d}x).$$

with respect to $\lambda_{|\mathcal{Y}}$, where $\lambda_{|\mathcal{Y}}$ is the restriction of λ onto Y. We assume that there exists a "true" parameter $\theta^* \in \Theta$ and let $\theta_1 \in \Theta$ be an arbitrary parameter. The general idea is to maximize p_θ instead of g_θ. Since the complete data is not given, the expected value under the previous estimate $\theta_k, k \in \mathbb{N}$ of the complete likelihood function given the observations $Y = y$ is maximized. For $k \in \mathbb{N}$ and $\theta_k \in \Theta$ the iteration of the EM algorithm is therefore given by

$$\theta_{k+1} \in \operatorname*{argmax}_{\theta \in \Theta} Q(\theta \mid \theta_k),$$

where

$$Q(\theta \mid \theta_k) := \mathbb{E}_{\theta_k}\left[\log(p_\theta(X)) \mid Y = y\right]. \tag{5.4}$$

Note that the starting parameter $\theta_0 \in \Theta$ can be chosen arbitrarily. Here, the expectation is taken with respect to the conditional density of X given Y, i.e.,

$$\mathbb{E}_{\theta_k}\left[X \mid Y = y\right] = \int\limits_{x \in \mathcal{X}} x \, \frac{p_{\theta_k}(x)}{g_{\theta_k}(y)} \, \mathrm{d}x.$$

We distinguish two steps:

E-step: Given $\theta_k \in \Theta$, determine $Q(\theta \mid \theta_k)$.

M-step: Choose $\theta_{k+1} \in \Theta$ to be any value in set $\operatorname*{argmax}_{\theta \in \Theta} Q(\theta \mid \theta_k)$.

The following Proposition verifies the idea.

Proposition 5.6. *Let $\ell(\cdot)$ be the log-likelihood function of Y and $(\theta_k)_{k \in \mathbb{N}}$ be an instance of the EM algorithm. Then, for all $k \in \mathbb{N}$ we have*

$$\ell(\theta_{k+1}) \geq \ell(\theta_k).$$

Proof. Let $\theta, \theta' \in \Theta$. Then

$$\begin{aligned}
\ell(\theta') &= \log g_{\theta'}(y) \\
&= \mathbb{E}_\theta\left[\log(g_{\theta'}(Y)) \mid Y = y\right] \\
&= \mathbb{E}_\theta\left[\log(p_{\theta'}(X)) \mid Y = y\right] - \mathbb{E}_\theta\left[\log(p_{\theta'}(X)) \mid Y = y\right] + \mathbb{E}_\theta\left[\log(g_{\theta'}(Y)) \mid Y = y\right] \\
&= Q(\theta' \mid \theta) - H(\theta' \mid \theta)
\end{aligned}$$

where

$$H(\theta' \mid \theta) = \mathbb{E}_\theta\left[\log\left(\frac{p_{\theta'}(X)}{g_{\theta'}(Y)}\right) \mid Y = y\right].$$

Jensen's inequality implies that

$$H(\theta \mid \theta_k) \leq H(\theta_n \mid \theta_k), \ \forall \theta \in \Theta, \tag{5.5}$$

which completes the proof.   □

Remark 5.7. *Since (5.5) holds, it follows that the log-likelihood increases in each step at least by*

$$Q(\theta_{n+1} \mid \theta_n) - Q(\theta_n \mid \theta_n).$$

Despite the property that we do not decrease the value of the likelihood function in any iteration, there is no guarantee that the EM-algorithm will converge to a global maximum. This is due to the fact that the likelihood function in general is multimodal. In fact, we have to make additional assumptions to ensure convergence to a local maximum of the likelihood. We define $\mathscr{M}$ to be the set of local maxima of ℓ and $\mathscr{S}$ as the set of saddle points of ℓ in the interior of Θ.

Theorem 5.8. *(Wu, 1983, Theorem 3) Suppose that Θ is compact and $Q(\cdot \mid \cdot)$ is continuous in both arguments, where Q is defined as in (5.4). If*

$$\max_{\theta \in \Theta} Q(\theta \mid \theta') > Q(\theta' \mid \theta'), \text{ for any } \theta' \in \mathscr{S} \setminus \mathscr{M} \tag{5.6}$$

Then all limit points of $(\theta_k)_{k \geq 1}$ of the EM algorithm are local maxima of ℓ and $\ell(\theta_k) \to \ell(\theta_0)$ as $k \to \infty$ for some local maximum $\theta_0 \in \mathscr{M}$.

Remark 5.9. *Condition 5.6 is satisfied for any density p_θ belonging to the class of standard exponential families.*

The Baum-Welch algorithm

The estimation of the parameters of the inhomogeneous hidden Markov model $(X_n, Z_n)_{n \in \mathbb{N}}$ can be regarded as a missing data problem. The incomplete data is the observation sequence $Z_1 = z_1, \ldots, Z_n = z_n$, while the complete data is the joint Markov chain $(X_1 = x_1, Z_1 = z_1), \ldots, (X_n = x_n, Z_n = z_n)$. For $t \in \mathbb{N}$ and $i, j \in S$ let $\xi_t(i, j) = \mathbb{P}_\theta^\nu(X_t = i, X_{t+1} = j \mid Z_1 = z_1, \ldots, Z_n = z_n)$ be the conditional probability of the states i and j at time t and $t + 1$ respectively conditioned on the observed sequence $z_1, \ldots, z_n$. The following proposition relates the conditional probabilities with the *forward* and *backward* variables.

Proposition 5.10. *Fix $\theta \in \Theta$ and $\nu \in \mathcal{P}(S)$. For any $i, j \in S$ and $t \in \{1, \ldots, n-1\}$ it holds that*

$$\xi_t(i, j) = \frac{\alpha_t(i) P_\theta(i, j) f_{\theta, t+1}(j, z_{t+1}) \beta_{t+1}(j)}{\sum\limits_{s_1=1}^{K} \sum\limits_{s_2=1}^{K} \alpha_t(s_1) P_\theta(s_1, s_2) f_{\theta, t+1}(s_2, z_{t+1}) \beta_{t+1}(s_2)}$$

Proof. See Appendix A.   □

For $\theta \in \Theta$ and $\nu \in \mathcal{P}(S)$ denote by $p_\theta^{\nu, X, Z}(x_1, z_1, \ldots, x_n, z_n)$ the likelihood of the complete data $(x_1, z_1), \ldots, (x_n, z_n)$. Note that

$$p_\theta(x_1, z_1, \ldots, x_n, z_n)^{\nu, X, Z} = \nu(x_1) f_{\theta,1}(x_1, z_1) \prod_{i=2}^{n} P_\theta(x_{i-1}, x_i) f_{\theta, i}(x_i, z_i).$$

Hence, for $\theta, \theta' \in \Theta$, the E-step reduces to

$$\begin{aligned}
&Q(\theta' \mid \theta) \\
&= \mathbb{E}_\theta^\nu \left[\log \left(p_{\theta'}^{\nu, X, Z}(X_1, Z_1, \ldots, X_n, Z_n) \right) \mid Z_1 = z_1, \ldots, Z_n = z_n \right] \\
&= \sum_{s_{1:n} \in S^n} \log \left(p_{\theta'}^{\nu, X, Z}(s_1, z_1, \ldots, s_n, z_n) \right) \mathbb{P}_\theta^\nu (X_1 = s_1, \ldots, X_n = s_n \mid Z_1 = z_1, \ldots, Z_n = z_n) \\
&= \sum_{s_{1:n} \in S^n} \left(\log(\nu(s_1)) + \sum_{i=1}^{n-1} \log(P_{\theta'}(s_i, s_{i+1})) + \sum_{i=1}^{n} \log(f_{\theta', i}(s_i, z_i)) \right) \mathbb{P}_\theta^\nu (X_{1:n} = s_{1:n} \mid Z_{1:n} = z_{1:n}) \\
&= \sum_{i=1}^{K} \log(\nu(i)) \gamma_1(i) + \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{t=1}^{n-1} \log(P_{\theta'}(i, j)) \xi_t(i, j) + \sum_{i=1}^{K} \sum_{t=1}^{n} \log(f_{\theta', t}(i, z_t)) \gamma_t(i), \quad (5.7)
\end{aligned}$$

where

$$\gamma_t(i) = \mathbb{P}_\theta^\nu (X_t = i \mid Z_1 = z_1, \ldots, Z_n = z_n) = \sum_{j=1}^{K} \xi_t(i, j), \quad i \in S, t \in \{1, \ldots, n\}.$$

Let be $k \in \mathbb{N}$ and $\theta_k \in \Theta \subset \mathbb{R}^d$. For $d_1, d_2 \in \mathbb{N}$ with $d_1 + d_2 K = d$ assume that we can decompose $\theta_k = (\rho_k, \phi_k^{(1)}, \ldots, \phi_k^{(K)})$ into a parameter $\rho_k \in \Theta_1 \subset \mathbb{R}^{d_1}$ and K parameters $\phi_k^{(1)}, \ldots, \phi_k^{(K)}$ with $\phi_k^{(i)} \in \Theta_2 \subset \mathbb{R}^{d_2}, i = 1, \ldots, K$. Furthermore, assume that P_{θ_k} is determined by ρ_k and $f_{\theta_k, i}(s, z)$ is determined by $\phi_k^{(s)}$, where $i \in \mathbb{N}, s \in S, z \in G$. Then the M-step can be decomposed into several separate maximization problems:

$$\rho_{k+1} \in \underset{\rho_k \in \Theta_1}{\arg\max} \sum_{t=1}^{n-1} \sum_{i=1}^{K} \sum_{j=1}^{K} \log(P_{\theta_k}(i, j)) \xi_n(i, j), \quad (5.8)$$

$$\phi_{k+1}^{(i)} \in \underset{\phi_k^{(i)} \in \Theta_2}{\arg\max} \sum_{t=1}^{n} \log(f_{\theta_k, t}(i, z_t)) \gamma_t(i), \quad i = 1, \ldots, K. \quad (5.9)$$

Furthermore, if $\rho_k = ((P_{\theta_k}(1, 1)), \ldots, (P_{\theta_k}(K-1, K-1)))^T$ the solution of the maximization

problem (5.8) is given by

$$(P_{\theta_{k+1}}(i,j)) = \frac{\sum_{t=1}^{n-1} \varepsilon_t(i,j)}{\sum_{t=1}^{n-1} \gamma_t(j)}, \quad i,j \in S.$$

The maximization in (5.9) depends on the density function $f_{\theta_k,t}$. In general, a closed-form solution is not guaranteed.

A forward algorithm for filtered Gaussian models

In this section we will neglect the additional inhomogeneous noise and propose a forward algorithm for filtered data. Assume that the conductance level recordings of an ion channel follows a Gaussian HMM, i.e., there exists an underlying Markov chain $X = (X_n)_{n\in\mathbb{N}}$ on a finite state space $S = \{1,\ldots,K\}$ governed by an irreducible transition matrix P_θ. The conductance level recordings $(\tilde{Y}_n)_{n\in\mathbb{N}}$ are given by

$$\tilde{Y}_n = \mu_\theta^{(X_n)} + \sigma_\theta^{(X_n)} V_n,$$

where $\mu \in \mathbb{R}^K, \sigma \in \mathbb{R}_+^K$ and $(V_i)_{i\in\mathbb{N}}$ are iid random variables with $V_1 \sim \mathcal{N}(0,1)$. Further we assume that $\theta \in \mathbb{R}^{(K-1)^2+2K}$ and

$$\theta = (P_\theta(1,1),\ldots,P_\theta(K-1,K-1),\mu_\theta^{(1)},\ldots,\mu_\theta^{(K)},(\sigma_\theta^{(1)})^2,\ldots,(\sigma_\theta^{(K)})^2)^T.$$

Ion channel recordings are usually filtered, which averages the conductance levels according to the filter coefficients, see Sigworth (1986). We will focus on the case where the filter $B = (B^{(0)},\ldots,B^{(b-1)})$ is discrete with finite length b such that

$$\sum_{j=0}^{b-1} B^{(j)} = 1.$$

Then the observed sequence $(Y_n)_{n\in\mathbb{N}}$ is modeled by

$$Y_n = \sum_{j=0}^{b-1} B^{(j)} \tilde{Y}_{n-j}.$$

For $n \in \mathbb{N}$ with $n \geq 2b-1$ we write $\mathbf{y}_{n-1} = (y_{n-1},\ldots,y_{n-b+1})$ and $\mathbf{x}_n = (x_n,\ldots,x_{n-2b+2})$ and similarly for $\mathbf{X}_n$ and $\mathbf{Y}_{n-1}$. Observe that conditioned on $\mathbf{X}_n = \mathbf{x}_n$, we have that $(Y_n, \mathbf{Y}_{n-1})$ is multivariate normally distributed with mean

$$\bar{\mu} = (\bar{\mu}^{(1)},\ldots,\bar{\mu}^{(b)}),$$

where

$$\bar{\mu}^{(i)} = \sum_{j=0}^{b-1} B^{(j)} \mu^{(\mathbf{x}_n^{(i+j)})}, \qquad i = 1, \dots, b$$

and covariance matrix

$$\Sigma^2 = \begin{pmatrix} \Sigma_{1,1}^2 & \Sigma_{1,2}^2 \\ \Sigma_{2,1}^2 & \Sigma_{2,2}^2 \end{pmatrix},$$

with $\Sigma_{1,1}^2 \in \mathbb{R}_+$, $\Sigma_{2,1}^2 \in \mathbb{R}^{b-1}$, $\Sigma_{1,2}^2 = (\Sigma_{2,1}^2)^T$ and $\Sigma_{2,2}^2 \in \mathbb{R}^{b-1 \times b-1}$. The covariance matrix Σ^2 is symmetric and the lower triangular entries are given by

$$(\Sigma^2)_{i,k} = \sum_{j=0}^{b-1-(i-k)} B^{(j)} B^{(j+(i-k))} \left(\sigma^{(\mathbf{x}_n^{(j+i+(i-k))})} \right)^2, \quad 1 \le k \le i \text{ with } i \le b.$$

It follows that

$$Y_n \mid \mathbf{Y}_{n-1} = \mathbf{y}, \mathbf{X}_n = \mathbf{x}_n \sim \mathcal{N}\left(\bar{\mu}^{(1)} + \Sigma_{1,2} \Sigma_{2,2}^{-1} (\mathbf{y} - (\bar{\mu}^{(2)}, \dots, \bar{\mu}^{(n)})), \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1} \right). \quad (5.10)$$

We see that the computation of the conditional likelihood of Y_n involves the $2b - 1$ previous states of the underlying Markov chain. This leads to computational costs in the E-step of nK^{4b-2}. Although there are procedures for filtered ion channel data, see for example Venkataramanan et al. (2000), Qin et al. (2000) or de Gunst and Schouten (2005), unfortunately none of these methods can be computed in suitable time. First, the number of data points is usually higher than 10^7. Second, the filter we deal with has at least 6 significant components.

Therefore we propose a modified *forward* algorithm which has computation cost in the E-step of nK^{2b-1}. The idea is based on the following observations in ion channel recordings. The filter coefficients decrease in time, i.e., $B^{(i)} > B^{(j)}$ for $i < j$. This implies that for any integer n, m with $n \ge m$, the influence of X_m and Y_m on Y_n decreases, if $n - m$ increases. Further, we observe that the probability that $X_n \ne X_{n-1}$ is smaller than 0.5. The basic idea is to replace $\mathbf{x}_n = (x_n, \dots, x_{n-b+1}, \dots, x_{n-2b+2}) \in S^{2b-1}$ by $\tilde{\mathbf{x}}_n = (x_n, \dots, x_{n-b+1}, \dots, x_{n-b+1}) \in S^{2b-1}$. Instead of using (5.10), we propose to use

$$Y_n \mid \mathbf{Y}_{n-1} = \mathbf{y}, \mathbf{X}_n = \mathbf{x}_n \sim \mathcal{N}\left(\tilde{\mu}^{(1)} + \tilde{\Sigma}_{1,2} \tilde{\Sigma}_{2,2}^{-1} (\mathbf{y}_{n-1} - (\tilde{\mu}^{(2)}, \dots, \tilde{\mu}^{(n)})), \tilde{\Sigma}_{1,1} - \tilde{\Sigma}_{1,2} \tilde{\Sigma}_{2,2}^{-1} \tilde{\Sigma}_{2,1} \right) \quad (5.11)$$

to compute the *forward* variables, where

$$\tilde{\mu}^{(i)} = \sum_{j=0}^{b-1-i} B^{(j)} \mu^{(\mathbf{x}_n^{(i+j)})} + \left(1 - \sum_{j=0}^{b-1-i} B^{(j)} \right) \mu^{(\mathbf{x}_n^{(b)})}, \qquad i = 1, \dots, b$$

and

$$(\tilde{\Sigma}^2)_{i,k} = \sum_{j=0}^{b-1-(i-k)-i} B^{(j)} B^{(j+(i-k))} \left(\sigma^{(\mathbf{x}_n^{(j+i+(i-k))})} \right)^2 + \sum_{j=b-1-(i-k)-i+1}^{b-1-(i-k)} B^{(j)} B^{(j+(i-k))} \left(\sigma^{(\mathbf{x}_n^{(b)})} \right)^2.$$

Remark 5.11. *If for all $s \in S$ we have*

$$P(s, s)^b > \max_{s_1, \dots, s_b} P(s, s_1) \prod_{i=2}^{b-1} P(s_{i-1}, s_{i-2}),$$

then we replace $\boldsymbol{x}_n \in S^{2b-1}$ in the proposed algorithm with the most likely sequence of states $\tilde{\boldsymbol{x}}_n \in S^{2b-1}$ such that the last b entries of $\tilde{\boldsymbol{x}}_n$ are equal. A backward algorithm based on this idea seems inappropriate, due to the replacing procedure. Therefore we use the computed forward variables to estimate the parameter.

Simulations and data analysis

In this section we will perform simulation of the models introduced in Section 3. We will perform maximum likelihood estimation and quasi-maximum likelihood estimation with the algorithms described in Section 5. Furthermore, we will analyze a data set from PorB recordings.

6.1 Poisson model

Recall the model from Section 3.1. First, we want to illustrate that the Baum-Welch algorithm as described in Section 5 can be used to obtain approximates of the MLE. To this end we set $\beta_n = 0$ for $n \in \mathbb{N}$ and therefore $Z_n = Y_n$ for all $n \in \mathbb{N}$. We will denote the resulting parameter of the BW-algorithm by $\theta_{\nu,n}^{\mathrm{ML}}$. Note that in this homogeneous HMM

$$
\mathbb{P}_{\theta^*}^{\pi} \left(\lim_{n \to \infty} \left| \theta_{\nu,n}^{\mathrm{ML}} - \theta^* \right| \right) = 1,
$$
$$
\lim_{n \to \infty} n^{1/2} (\theta_{\nu,n}^{\mathrm{ML}} - \theta^*) \to \mathcal{N}(0, F^{-1}), \tag{6.1}
$$

where

$$
F = \lim_{n \to \infty} \frac{1}{n} \mathrm{E}_{\theta}^{\pi} \left[\left(\frac{\partial}{\partial \theta^*} \log q_{\theta^*}^{\pi}(Y_1, \ldots, Y_n) \right) \left(\frac{\partial}{\partial \theta^*} \log q_{\theta^*}^{\pi}(Y_1, \ldots, Y_n) \right)^T \right].
$$

We refer to F as the Fisher Information. Unfortunately, there exists no closed-form formula to compute F. Therefore we use a Monte Carlo simulation with $t = 10^3$ trials and $n = 10^5$ observations to compute F. We simulate under the following setting. Let $K = 2$, $\theta^* = (0.6, 0.2, 10, 25)$, $P_{\theta^*}(1, 1) = 0.6$, $P_{\theta^*}(1, 1) = 0.2$ and $\lambda = (10, 25)$. Figure 6.1 shows a representative trajectory of $(Y_n)_{n \in \mathbb{N}}$.

The Monte-Carlo simulation leads to

$$
F = \begin{pmatrix}
1.21 & 0.22 & -0.015 & 0 \\
0.22 & 3.76 & 0.03 & -0.02 \\
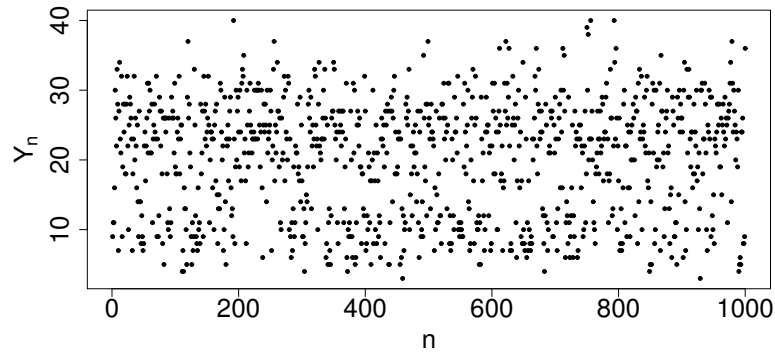-0.01 & -0.03 & 0.03 & 0 \\
0 & -0.02 & 0 & 0.03
\end{pmatrix}
$$

Figure 6.1: Exemplary trajectory of model 3.1 with 10^3 observations and $K = 2$, $\theta^* = (0.6, 0.2, 10, 25)$, $P_{\theta^*}(1, 1) = 0.6$, $P_{\theta^*}(2, 1) = 0.2$, $\lambda = (10, 25)$, $\nu = (1/2, 1/2)$ and $\beta_n = 0$ for $n \in \mathbb{N}$.

and

$$
F^{-1} = \begin{pmatrix}
0.84 & -0.05 & 0.44 & 0.04 \\
-0.05 & 0.27 & 0.34 & 0.26 \\
0.44 & 0.34 & 41.45 & 8.29 \\
0.04 & 0.26 & 8.29 & 41.32
\end{pmatrix}.
$$

For $j \in \{1, \ldots, t\}$ denote by $\theta_{\nu,n}^{\mathrm{ML}}(j)$ the ML estimator of θ^* in the j-th trial computed by the Baum-Welch algorithm. Further, for $k \in \{1, \ldots, 4\}$ let $\mu^{\mathrm{ML}}(k)$ be the sample mean and $\sigma^{\mathrm{ML}}(k)$ be the sample variance of the k-th component of the scaled estimators, i.e.,

$$
\mu^{\mathrm{ML}}(k) = t^{-1} \sum_{j=1}^{t} n^{-1/2} \left((\theta_{\nu,n}^{\mathrm{ML}}(j))^{(k)} - \theta^* \right)
$$

and

$$
\sigma^{\mathrm{ML}}(k) = (t-1)^{-1} \sum_{j=1}^{t} n^{-1/2} \left((\theta_{\nu,n}^{\mathrm{ML}}(j))^{(k)} - \mu^{\mathrm{ML}}(k) \right)^2.
$$

For $k = 1, \ldots, 4$ Table 6.1 compares $\mu^{\mathrm{ML}}(k)$ and $\sigma^{\mathrm{ML}}(k)$ with the theoretical values from equation (6.1). We observe that the BW-algorithm performs very well in the sense that it reaches the theoretical boundaries of the MLE.

Parameter component	$\left\|\mu^{\mathrm{ML}}(k)\right\|$	$F^{-1}(k, k)$	$\left\|F^{-1}(k, k) - \sigma^{\mathrm{ML}}(k)\right\|$
$P_{\theta^*}(1, 1)$	0.02	0.84	0.13
$P_{\theta^*}(2, 1)$	0.01	0.27	0
$\lambda_{\theta^*}^{(1)}$	0.12	41.45	1.02
$\lambda_{\theta^*}^{(2)}$	0.15	41.32	2.56

Table 6.1: Component-wise comparison of the theoretical mean and theoretical variance of $\lim_{n \to \infty} n^{1/2}(\theta_{\nu,n}^{\mathrm{ML}} - \theta^*)$ obtained by Monte Carlo simulation with the sample mean μ^{ML} and sample variance σ^{ML} in the Poisson model.

Now we consider an inhomogeneous HMM with inhomogeneous intensity $\beta_n = 10n^{-1.1}, n \in \mathbb{N}$. We leave the other parameters unchanged and compare the performance of $\theta_{\nu,n}^{\text{ML}}$ and $\theta_{\nu,n}^{\text{QML}}$ in Figure 6.2. We see that both estimators converge to θ^*. Naturally, $\theta_{\nu,n}^{\text{ML}}$ outperforms $\theta_{\nu,n}^{\text{QML}}$, since the inhomogenity is explicitly modeled.
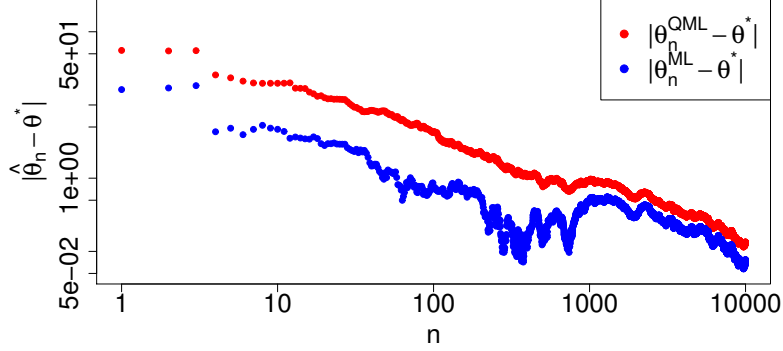

Figure 6.2: Euclidean distance between $\theta_{\nu,n}^{\text{QML}}$ and θ^* and between $\theta_{\nu,n}^{\text{ML}}$ and θ^* in the above described Poisson model with $\beta_n = 10n^{-1.1}, n \in \mathbb{N}$.

In the following we analyze the asymptotic behavior of $n^{1/2}(\theta_{\nu,n}^{\text{ML}} - \theta^*)$ and $n^{1/2}(\theta_{\nu,n}^{\text{QML}} - \theta^*)$. To this end we generate $t = 10^3$ trajectories of the above described model with $n = 10^5$ observations. Figure 6.3 and Figure 6.4 show representative sequences of estimates for $P_{\theta^*}(1,1)$ and $\lambda_{\theta^*}^{(1)}$, respectively. We observe that the absolute values of both estimators are almost equal.
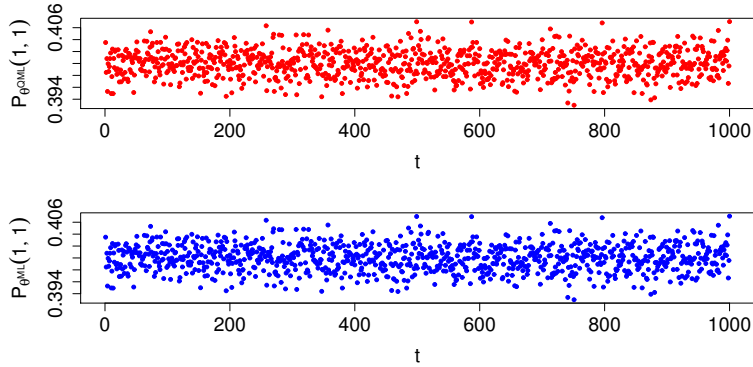

Figure 6.3: Exemplary sequence of $P_{\theta^{\text{QML}}}(1,1)$ (top) and $P_{\theta^{\text{ML}}}(1,1)$ (bottom) in the inhomogeneous Poisson model with 10^3 trajectories.

Recall the definitions of $G_{n,\text{QML}}$, $G_{n,\text{ML}}$ F_{QML} and $F_{n,\text{ML}}$ from Section 2. We compute $G_{n,\text{QML}}$, $G_{n,\text{ML}}$ $F_{n,\text{QML}}$ and $F_{n,\text{ML}}$ numerically via a Monte Carlo simulation and observe that all quantities converges to F. It follows that the $\theta_{\nu,n}^{\text{ML}}$ and $\theta_{\nu,n}^{\text{QML}}$ in the inhomogeneous model have the same Cramér-Rao bound as maximum likelihood estimator in the homogeneous case. For $k \in \{1, \dots, 4\}$ define $\mu^{\text{QML}}(k)$ and $\sigma^{\text{QML}}(k)$ analogously to $\mu^{\text{ML}}(k)$ and $\sigma^{\text{ML}}(k)$. For $k \in \{1, \dots, 4\}$ we compute the empirical means $\mu^{\text{ML}}(k)$, $\mu^{\text{QML}}(k)$ and the empirical variances $\sigma^{\text{ML}}(k)$, $\sigma^{\text{QML}}(k)$ and compare them with $F^{-1}(k,k)$. Table 6.2 illustrates that $\theta_{\nu,n}^{\text{QML}}$ and $\theta_{\nu,n}^{\text{ML}}$ are asymptotically optimal in the
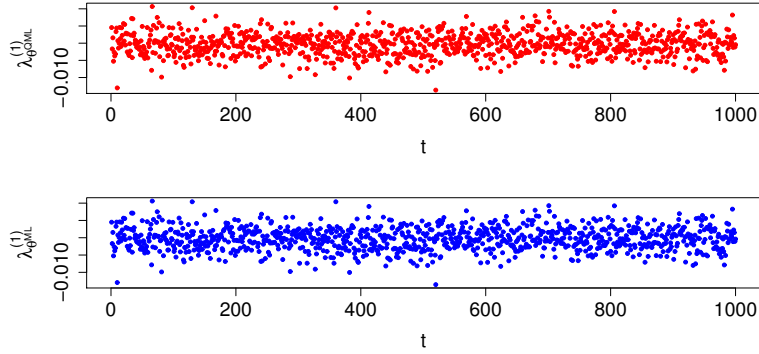
Figure 6.4: Exemplary sequence of $\lambda^{(1)}_{\theta\,\mathrm{QML}}$ (top) and $\lambda^{(1)}_{\theta\,\mathrm{ML}}$ (bottom) in the inhomogeneous Poisson model with 10^3 trajectories.

sense that they reach the variance boundaries from the homogeneous case.

Parameter component	$\left|\mu^{\mathrm{QML}}(k)\right|$	$\left|\mu^{\mathrm{ML}}(k)\right|$	$\left|\sigma^{\mathrm{QML}}(k) - F^{-1}(k,k)\right|$	$\left|\sigma^{\mathrm{ML}}(k) - F^{-1}(k,k)\right|$
$P_{\theta^*}(1,1)$	0.02	0.02	0.05	0.05
$P_{\theta^*}(2,1)$	0	0	0.01	0.01
$\lambda^{(1)}_{\theta^*}$	0.30	0.27	0.87	0.91
$\lambda^{(2)}_{\theta^*}$	0.09	0.01	2.40	2.46

Table 6.2: Component-wise comparison of the theoretical mean and theoretical variance of $\lim_{n\to\infty} n^{1/2}(\theta^{\mathrm{ML}}_{\nu,n} - \theta^*)$ obtained by Monte Carlo simulation with the sample means μ^{ML}, μ^{QML} and sample variances σ^{ML}, σ^{QML} in the Poisson model with $\beta_n = 10n^{-1.1}$, $n \in \mathbb{N}$.

6.2   Gaussian model

Recall the model from Section 3.2. Again, we use the Baum-Welch algorithm as described in Section 5 to compute the MLE. Similarly to the previous section we set $\beta_n = 0$ for all $n \in \mathbb{N}$ for the moment. Furthermore we choose $M = 1$, $K = 2$, $n = 10^5$, $\theta^* = (0.1, 0.5, 10, 20, 5, 5)$, $\mathbb{P}_{\theta^*}(1,1) = 0.1$, $\mathbb{P}_{\theta^*}(2,1) = 0.5$, $\mu_{\theta^*} = (10, 20)$ and $\sigma^2_{\theta^*} = (5,5)$. The Fisher information F in the homogeneous model has been computed numerically via Monte Carlo simulation with $t = 10^3$ trials. The inverse of F is given by

$$
F^{-1} = \begin{pmatrix}
0.28 & 0.02 & 0.21 & 0.03 & 0.76 & -0.46 \\
0.02 & 0.46 & 0.33 & 0.18 & 1.08 & -0.60 \\
0.21 & 0.33 & 16.85 & 1.74 & 11.98 & -5.85 \\
0.03 & 0.18 & 1.74 & 9.57 & 6.31 & -3.37 \\
0.76 & 1.08 & 11.98 & 6.31 & 191.59 & -21.97 \\
-0.46 & -0.60 & -5.85 & -3.37 & -21.97 & 91.59
\end{pmatrix}.
$$

For $j \in \{1, \ldots, t\}$ denote by $\theta^{\mathrm{ML}}_{v,n}(j)$ the ML estimator of θ^* in the j-th trial computed by the Baum-Welch algorithm. Similarly to Section 6.1 we define the sample mean and the sample variance of a sequence of estimators. For $k \in \{1, \ldots, 6\}$ let $\mu^{\mathrm{ML}}(k)$ be the sample mean and $\sigma^{\mathrm{ML}}(k)$ be the sample variance of the k-th component of the scaled estimators, i.e.,

$$\mu^{\mathrm{ML}}(k) = t^{-1} \sum_{j=1}^{t} n^{-1/2} \left((\theta^{\mathrm{ML}}_{v,n}(j))^{(k)} - \theta^* \right)$$

and

$$\sigma^{\mathrm{ML}}(k) = (t-1)^{-1} \sum_{j=1}^{t} n^{-1/2} \left((\theta^{\mathrm{ML}}_{v,n}(j))^{(k)} - \mu^{\mathrm{ML}}(k) \right)^2 .$$

For $k = 1, \ldots, 6$ Table 6.3 compares $\mu^{\mathrm{ML}}(k)$ and $\sigma^{\mathrm{ML}}(k)$ with the theoretical mean and variance of $\lim_{n \to \infty}(\theta^{\mathrm{ML}} - \theta^*)$. Similarly to the Poisson model in the previous section, the performance of θ^{ML} is very close the theoretical boundaries.

Parameter component	$\left|\mu^{\mathrm{ML}}(k)\right|$	$F^{-1}(k,k)$	$\left|\sigma^{\mathrm{ML}}(k) - F^{-1}(k,k)\right|$
$P_{\theta^*}(1,1)$	0.02	0.28	0.02
$P_{\theta^*}(2,1)$	0	0.46	0.06
$\mu^{(1)}_{\theta^*}$	0.16	16.85	0.31
$\mu^{(2)}_{\theta^*}$	0.02	9.57	1.64
$(\sigma^2_{\theta^*})^{(2)}$	0.15	191.59	8.93
$(\sigma^2_{\theta^*})^{(2)}$	0.19	91.59	14.90

Table 6.3: Component-wise comparison of the theoretical mean and theoretical variance of $\lim_{n \to \infty} n^{1/2}(\theta^{\mathrm{ML}}_{v,n} - \theta^*)$ obtained by Monte Carlo simulation with the sample mean μ^{ML} and sample variance σ^{ML} in the linear Gaussian model.

In the following we will focus on the inhomogeneous case by setting $\beta^2_n = 40n^{-1}$ for $n \in \mathbb{N}$. Figures 6.5 shows an representative trajectory of the inhomogeneous model.
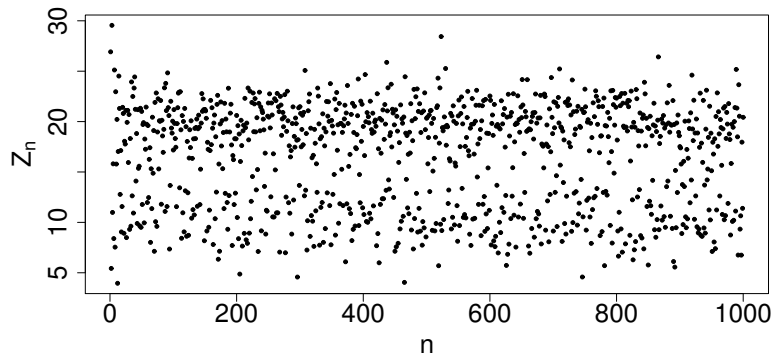

Figure 6.5: Exemplary trajectory of 10^3 observations of the inhomogeneous normal with $M = 1$, $K = 2$, $n = 10^5$, $\theta^* = (0.1, 0.5, 10, 20, 5, 5)$, $\mathbb{P}_{\theta^*}(1,1) = 0.1$, $\mathbb{P}_{\theta^*}(2,1) = 0.5$, $\mu_{\theta^*} = (10, 20)$, $\sigma^2_{\theta^*} = (5, 5)$ and $\beta_i = 40i^{-1}, \in \in \mathbb{N}$.

In Figure 6.6 we compare the performance of $\theta_{v,n}^{\mathrm{ML}}$ and $\theta_{v,n}^{\mathrm{QML}}$. Surprisingly, the performance of $\theta_{v,n}^{\mathrm{QML}}$ seems to be slightly better for small $n \in \mathbb{N}$. One reason for this could be that in the M-step for $\theta_{v,n}^{\mathrm{ML}}$, no closed-form solution for the variance is available. Therefore we use an approximate maximal value.
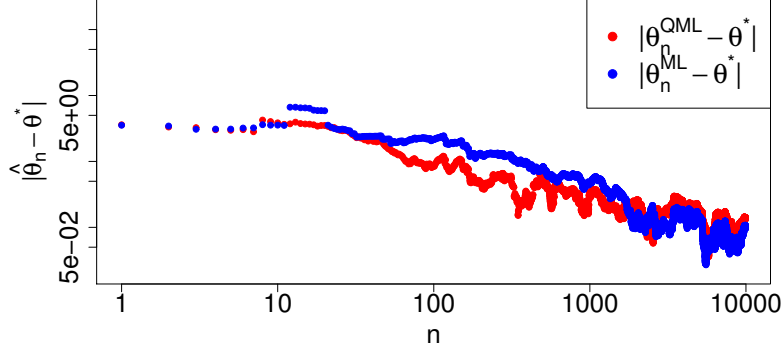

Figure 6.6: Euclidean distance between $\theta_{v,n}^{\mathrm{QML}}$ and θ^* and Euclidean distance between $\theta_{v,n}^{\mathrm{ML}}$ and θ^* in the normal model.

The asymptotic behavior of $n^{1/2}(\theta_{v,n}^{\mathrm{ML}} - \theta^*)$ and $n^{1/2}(\theta_{v,n}^{\mathrm{QML}} - \theta^*)$ is similar to the Poisson model, since again $G_{n,QML}$, $G_{n,ML}$ $F_{n,QML}$ and $F_{n,ML}$ converges to F. Note that for $i \in \{1, \ldots, 6\}$ an asymptotic confidence interval for $(\theta^*)^{(i)}$ with error rate $\alpha \in (0, 1)$ is given by

$$\left[(\theta^{\mathrm{ML}})^{(i)} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{nF_{n,ML}^{-1}(i,i)}}, (\theta^{\mathrm{ML}})^{(i)} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{nF_{n,ML}^{-1}(i,i)}} \right]$$

or

$$\left[(\theta^{\mathrm{QML}})^{(i)} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{nF_{n,QML}^{-1}(i,i)}}, (\theta^{\mathrm{QML}})^{(i)} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{nF_{n,QML}^{-1}(i,i)}} \right],$$

where z_q is the q-quantile of a standard normal distribution.

6.2.1 Slowly decreasing inhomogeneous noise

In this Section we investigate the effects on the asymptotic behavior of the MLE and the quasi-MLE in the Gaussian linear model from Section 3.2, if the inhomogeneous noise is slowly decreasing. To be precise, let $\beta_n^2 = \tau n^{-1/2}, \tau \in R_+, n \in \mathbb{N}$. Furthermore, we chose $M = 1$, $K = 1$, $\theta^* = (0, 1)$, $\mu_{\theta^*} = 0$ and $\sigma_{\theta^*} = 1$. Since $K = 1$, we have that $(Z_n)_{n \in \mathbb{N}}$ is a sequence of independent, normally distributed random variables with mean zero and variance $1 + \beta_n^2$. In the following we

will suppress the first argument of f_{θ^*}. Note that

$$
S_{n,\text{QML}}(\theta) = \sum_{i=2}^{n} \frac{\partial}{\partial\theta} \log f_\theta(Z_i)
$$

$$
= \left(\sum_{i=2}^{n} \frac{Z_i - \mu_\theta}{2\sigma_\theta^2}, \sum_{i=2}^{n} \frac{(Z_i - \mu_\theta)^2}{2(\sigma_\theta^2)^2} - \frac{1}{2\sigma_\theta^2} \right)^T.
$$

It follows that assumption (CLT1) is fulfilled if

$$
\lim_{n\to\infty} n^{-1/2} \sum_{i=1}^{n} \beta_i^2 = 0.
$$

However, we have

$$
\lim_{n\to\infty} n^{-1/2} \sum_{i=1}^{n} \beta_i^2 = 2\tau.
$$

Remark 6.1. *A similar calculation shows for the Poisson model from Section 3.1 with K = 1 and inhomogeneous intensity β_i for $i \in \mathbb{N}$ that condition (2.16) holds whenever*

$$
\beta_n = o(n^{-1/2}).
$$

We can still find an asymptotic law for $n^{1/2}(\theta_n^{\text{QML}} - \theta)$. To this end note that

$$
\lim_{n\to\infty} F_{n,\text{QML}} = \lim_{n\to\infty} n^{-1} \mathbb{E}_{\theta^*}\left[-\frac{\partial}{\partial\theta} S_{n,\text{QML}}(\theta^*) \right]
$$

$$
= \lim_{n\to\infty} n^{-1} \mathbb{E}_{\theta^*} \begin{pmatrix} \sum_{i=2}^{n} \frac{1}{2\sigma_{\theta^*}^2} & \sum_{=2}^{n} \frac{Z_i - \mu}{2(\sigma_{\theta^*}^2)^2} \\ \sum_{i=2}^{n} \frac{Z_i - \mu}{2(\sigma_{\theta^*}^2)^2} & \sum_{i=2}^{n} \frac{(Z_i - \mu)^2}{(\sigma_{\theta^*}^2)^3} - \frac{1}{2(\sigma_{\theta^*}^2)^2} \end{pmatrix}
$$

$$
= \begin{pmatrix} \frac{1}{2\sigma_{\theta^*}^2} & 0 \\ 0 & \frac{1}{2(\sigma_{\theta^*}^2)^2} + \lim_{n\to\infty} n^{-1} \sum_{i=2}^{n} \frac{\beta_i}{(\sigma_{\theta^*}^2)^3} \end{pmatrix}
$$

$$
= \begin{pmatrix} \frac{1}{2\sigma_{\theta^*}^2} & 0 \\ 0 & \frac{1}{2(\sigma_{\theta^*}^2)^2} \end{pmatrix}
$$

and similarly

$$
\lim_{n\to\infty} G_{n,\text{QML}} = \lim_{n\to\infty} n^{-1} \text{Var}_{\theta^*}(S_{n,\text{QML}}(\theta^*))
$$

$$
= \begin{pmatrix} \frac{1}{\sigma_{\theta^*}^2} & 0 \\ 0 & \frac{1}{2(\sigma_{\theta^*}^2)^2} \end{pmatrix}.
$$

It follows that

$$n^{1/2}(\theta_{\nu,n}^{\mathrm{QML}} - \theta) \overset{D}{\to} \mathcal{N}\left(\begin{pmatrix} 0 \\ 2\tau \end{pmatrix}, \begin{pmatrix} 2\sigma_{\theta^*}^2 & 0 \\ 0 & 2(\sigma_{\theta^*}^2)^2 \end{pmatrix}\right),$$

as $n \to \infty$. We want to stress, that $\theta_{\nu,n}^{\mathrm{QML}}$ is still strongly consistency. Figure 6.7 illustrates an exemplary trajectory of $\left|\theta_{\nu,n}^{\mathrm{QML}} - \theta^*\right|$.
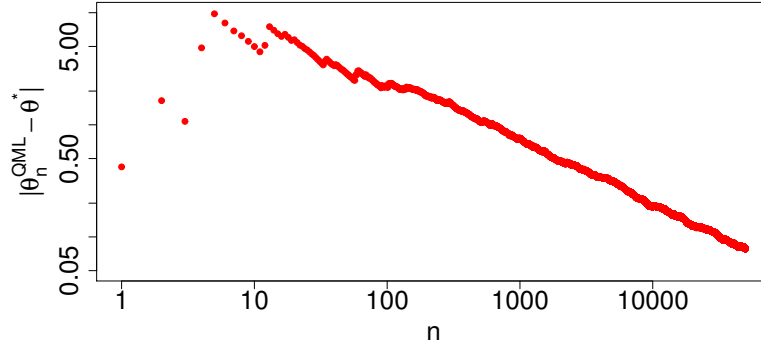

Figure 6.7: Representative trajectory of $\left|\theta_{\nu,n}^{\mathrm{QML}} - \theta^*\right|$ with $\tau = 10$.

An asymptotic confidence interval for $(\theta^*)^{(2)}$ is given by

$$I_{\tau,\sigma^2,n} = \left[(\theta_{\nu,n}^{\mathrm{QML}})^{(2)} - \frac{2\tau + z_{1-\frac{\alpha}{2}}}{(\theta_{\nu,n}^{\mathrm{QML}})^{(2)}\sqrt{2n}}, (\theta_{\nu,n}^{\mathrm{QML}})^{(2)} + \frac{2\tau - z_{1-\frac{\alpha}{2}}}{(\theta_{\nu,n}^{\mathrm{QML}})^{(2)}\sqrt{2n}} \right].$$

We want to compare the quality of $I_{\tau,\sigma^2,n}$ and the naive confidence interval $I_{\sigma^2,n}$, where

$$I_{\sigma^2,n} = \left[(\theta_{\nu,n}^{\mathrm{QML}})^{(2)} - \frac{z_{1-\frac{\alpha}{2}}}{(\theta_{\nu,n}^{\mathrm{QML}})^{(2)}\sqrt{2n}}, (\theta_{\nu,n}^{\mathrm{QML}})^{(2)} + \frac{z_{1-\frac{\alpha}{2}}}{(\theta_{\nu,n}^{\mathrm{QML}})^{(2)}\sqrt{2n}} \right].$$

To this end, we define the success rate $s_{\sigma_{\theta^*}^2}(I)$ of an interval $I = [a,b], a, b \in \mathbb{R}$ to be the relative frequency of successes, where a success is the event $\sigma_{\theta^*}^2 \in I$. Figure 6.8 shows the success rates of $I_{\tau,\sigma^2,n}$ and $I_{\sigma^2,n}$ with 10^3 trials as a function of the number observations for $\tau = 1$.

6.2.2   Filtered Gaussian model

In this section we want to simulate from the scenario we find in ion channel recordings. Note that the current recordings are filtered by a 4-pole Bessel filter B_{cont} with sampling rate 10^4 and cutoff frequency 10^3. Figures 6.9 shows its kernel function k. Therefore, for a time-continuous signal $(Y_t)_{t \in \mathbb{R}}$ the filtered signal $(W_t)_{t \in \mathbb{R}}$ is given by the convolution of W_t and B_{cont}, i.e.,

$$W_t = Y_t * B_{\mathrm{cont}} := \int_{-\infty}^{\infty} Y_{t-s} k(s) \mathrm{d}(s) = \int_{0}^{\infty} Y_{t-s} k(s) \mathrm{d}(s).$$
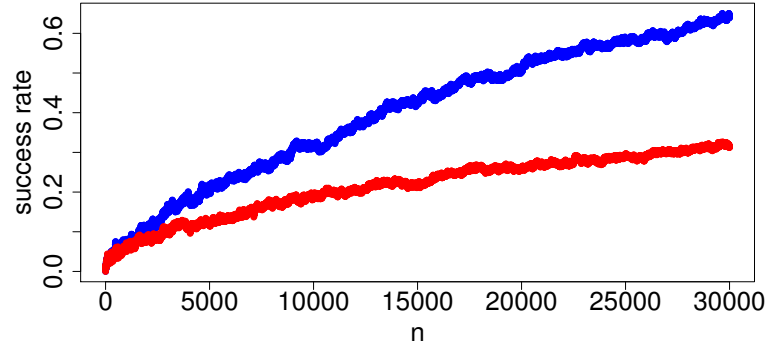
Figure 6.8: Success rate of $I_{\tau,\sigma^2,n}$ (blue) and $I_{\sigma^2,n}$ (red) for $\beta_n = n^{-1/2}, n \in \mathbb{N}$.
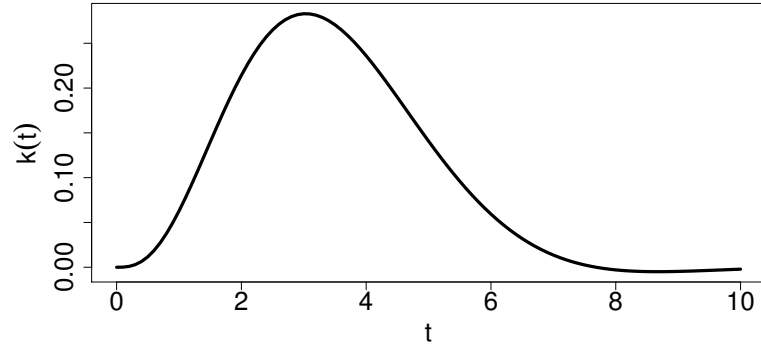

Figure 6.9: Kernel function k of a 4-pole Bessel filter with sampling rate 10^4 and cutoff frequency 10^3.

Since we observe a time-discrete process $(W_n)_{n\in\mathbb{N}}$, we approximate the filtered data by

$$W_n \approx \sum_{i=0}^{b-1} Y_{n-i} b_i,$$

where $b = 8$,

$$b_0 = \int_0^{0.5} k(s)\mathrm{d}(s)$$

and

$$b_i = \int_{i-0.5}^{i+0.5} k(s)\mathrm{d}(s), \quad i = 1,\dots,7.$$

The resulting discrete filter is given by $B = (0.002, 0.067, 0.21, 0.276, 0.232, 0.140, 0.060, 0.015)^T$. Assume now we are in the setting of the linear Gaussian model from Section 3.2 with $M = 1$ and $K = 2$, but instead of

$$Z_n = Y_n + \varepsilon_n,$$

we observe

$$Z_n = W_n + \varepsilon_n,$$

where

$$W_n = \sum_{i=0}^{b-1} Y_{n-i} b_i.$$

Furthermore, $\mu_{\theta^*} \in \mathbb{R}^2$ and $\sigma_{\theta^*}^2 \in \mathbb{R}_+^2$ are assumed to be known. This is a reasonable assumption in ion channel recordings, since due to the long-term persistence in each state these parameters can be estimated very well in advance. For the simulation study we assume that $\theta^* = (0.99, 0.4)$, $P_{\theta^*}(1, 1) = 0.99$, $P_{\theta^*}(2, 1) = 0.4$ and $\beta_n = 0.2n^{-1}$ for $n \in \mathbb{N}$. Furthermore we set $\mu_{\theta^*} = (2, 1)$ and $\sigma_{\theta^*}^2 = (0.1, 0.1)$. Figure 6.10 illustrates an exemplary trajectory of $(Z_n)_{n \in \mathbb{N}}$ together with a typical blockage event. We simulate $t = 10^3$ trajectories of $n = 10^6$ observations and estimated θ^* using
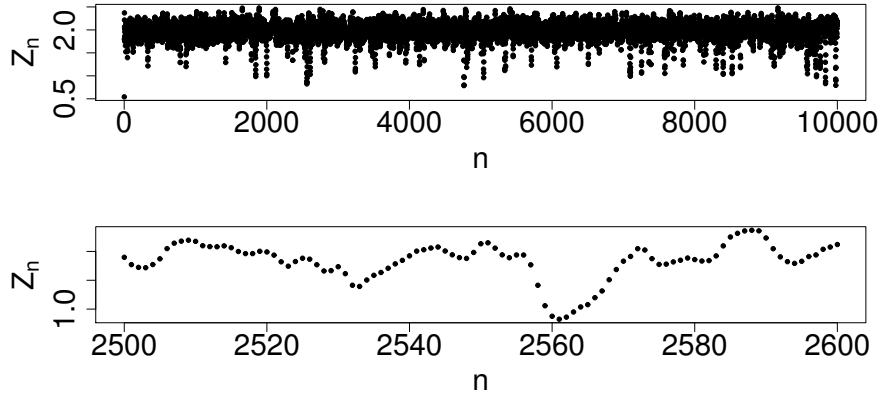

Figure 6.10: Exemplary trajectory of $(Z_n)_{n \in \mathbb{N}}$ (top) and blockage event (bottom).

the *forward* algorithm described in Section 5.2. The averaged estimated parameter θ_{av}^{QML} is given by $\theta_{av}^{\text{QML}} = (0.989, 0.398)$. This shows that the *forward* performs well.

6.3   Ion channel recordings

In this section we apply to the *forward* algorithm from Section 5.2 to ion channel recordings and present our results. The results concerning experiments with constant voltage can be found in Bartsch et al. (2017).

Against the background of multidrug-resistant bacteria we explored together with the Steinem lab (Institute of Organic and Biomolecular Chemistry, University of Göttingen) and other external collaborators the interaction of the antibiotic ampicillin with ion channels of the bacterial porin PorB. The broad-spectrum antibiotic irreversibly binds to and inhibits the activity of the transpeptidase enzyme, which occurs exclusively in bacteria. This inhibits the cell division of the bacteria and eventually leads to bactericide, see Acred et al. (1962). A potential source of antibiotic resistance is preventing the antibiotic to pass trough the outer bacterial membrane, see

Delcour (2009).

We studied the outer membrane porin PorB from Neisseria meningitidis, a pathogenic bacterium in the human nose and throat region. Two types of porins have been compared, a wild type and the mutant G103K. Cells with this mutation are suspected to be more likely resistant to antibiotics, see Oppenheim (1997) ansOlesky et al. (2002)).

Patch-clamp experiments were performed for a quantitative characterization of the interaction of the two types with ampicillin. The measurements are performed in the Steinem lab and use planar black lipid membranes (BLMs). When an ampicillin molecule binds to the pore it blocks the ion flow temporarily and, hence, this event can be detected by a conductance loss. Note that it cannot be decided whether an ampicillin molecule really passes trough the channel or only enters the channels but leaves to the same side. Single channel recordings using solvent-free bilayers at the Port-a-Patch were used to explore the conductivity of the wild type and the mutant without presence of ampicillin. We refer to Bartsch et al. (2017) for a deeper insight in the biological and medical background, as well as for the interpretations of the results.

6.3.1   Ion channel recordings with constant voltage

Ion channel recording can be sampled at frequencies ranging from 1 to 100 kHz. The gating events occur usually on much smaller times scales, ranging from 1 ns to 100 ns. Hence, channel recordings have the appearance of abrupt random changes, see Hamill et al. (1981). Consequently, the conductance level of a channel is modeled by a piecewise constant signal

$$Y_t = \sum_{j=1}^{K} \mu^{(j)} \mathbb{1}_{\{j\}}(X_t),$$

where $t > 0$ denotes the physical time. The unknown state of the channel is denoted by $(X_t)_{t \in \mathbb{R}_+}, X_t \in \{1, \ldots, K\}$. The unknown conductance levels are denoted by $\mu^{(1)}, \ldots, \mu^{(K)}$, where each level corresponds to one state. We assume that $(X_t)_{t \in \mathbb{R}}$ is a time-homogeneous Markov chain.

The very small conductance of a single channel, typically in the range of picosiemens up to few nanosiemens, requires sophisticated electronic recordings devices, including one or several amplifiers, see Devices (2008). To stay in the transmission range of the amplifier, high frequent noise components, e.g., caused by shot noise, are attenuated by convolving the recordings with an analogue lowpass filter. Typically, a four, six or eight pole lowpass Bessel filter is integrated in the hardware of the technical measurement device. Finally, the recorded currents are digitized equidistantly with sample rate f_s and divided by the applied constant voltage. Additionally, we assume that the signal $(Y_t)_{t \in \mathbb{R}_+}$ is perturbed by Gaussian white noise $(v_t)_{t \in \mathbb{R}_+}$. Thus, the recorded observations $W_1, \ldots, W_n$ are the filtered perturbed conductivity levels at equidistant time points $t_i = i/f_s$ for $i = 1, \ldots, n$ with an analogue lowpass filter having the kernel function $k, k : \mathbb{R} \to \mathbb{R}_+,$

of the Bessel filter, i.e.,

$$W_i = (k * (s + \sigma v))) (t_i) = \int_{-\infty}^{\infty} k(t_i - u)(Y_u + \sigma v_u) \mathrm{d}u, \quad i = 1, \ldots, n. \tag{6.2}$$

Here, $\sigma > 0$ denotes the standard deviation of the states and is assumed to be equal for all states. All of these measurements are recorded at sampling rate 50 kHz and were filtered with a four-pole Bessel lowpass filter with cutoff frequency 5 kHz, resulting in a normalized cutoff frequency of 0.1. As described in Section 6.2.2, we approximate the convolved observation by

$$W_n \approx \sum_{i=0}^{b-1} Y_{n-i} b_i,$$

where

$$b_0 = \int_0^{0.5} k(s) \mathrm{d}(s)$$

and

$$b_i = \int_{i-0.5}^{i+0.5} k(s) \mathrm{d}(s), \quad i = 1, \ldots, 7.$$

Furthermore, we assume that $(X_n)_{n \in \mathbb{N}}$ is a Markov chain with irreducible transition kernel P_{θ^*}. For the wild type as well as for the mutant G103K four measurements with 1 mM ampicillin concentration and at different applied voltage levels of $40, 60, 80, 100$ and 120 mV were recorded. Additionally, for both proteins the ampicillin was added in steps to obtain measurements with different ampicillin concentrations of $0.1, 0.2, 0.4, 0.6, 0.8$ and 1 mM at 80 mV. In each measurement the recordings last at least five minutes and, hence, at least 3 million were available. Figure 6.11 shows a representative recording of the wild type.

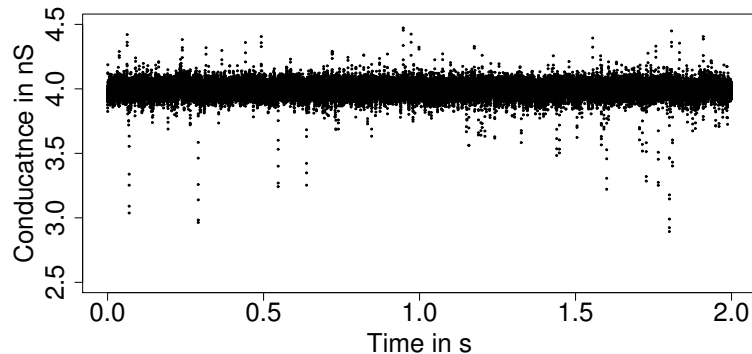Before we estimated the transition probabilities with the *forward*-algorithm, we do several

Figure 6.11: Observations of a representative conductance time series of 2 seconds of PorB wild type with 1 mM ampicillin recorded by the patch clamp technique using BLMs at 80 mV.

pre-processing steps. Data cleansing was necessary due to base line fluctuation and the presence

of outliers. To this end, we used the *JULES* procedure (Pein et al. (2017a)) to detect outliers and changes in the conductivity caused by the apparatus. In another preprocessing step we estimated the amplitudes of a blockage event. The estimated amplitudes are on average 1.19 nS for the wild type and 0.81 nS for the mutant. The estimated transition probabilities were used to determine the most likely sequence of states by the Viterbi algorithm, see Viterbi (1967). Then we used the idealization to compute the average blockage frequency and average residence time. In Figure 6.12 and 6.13 we compare our results with the findings of *JULES*. We stress that the averaged residence times and frequencies are very close to each other for all measurements.
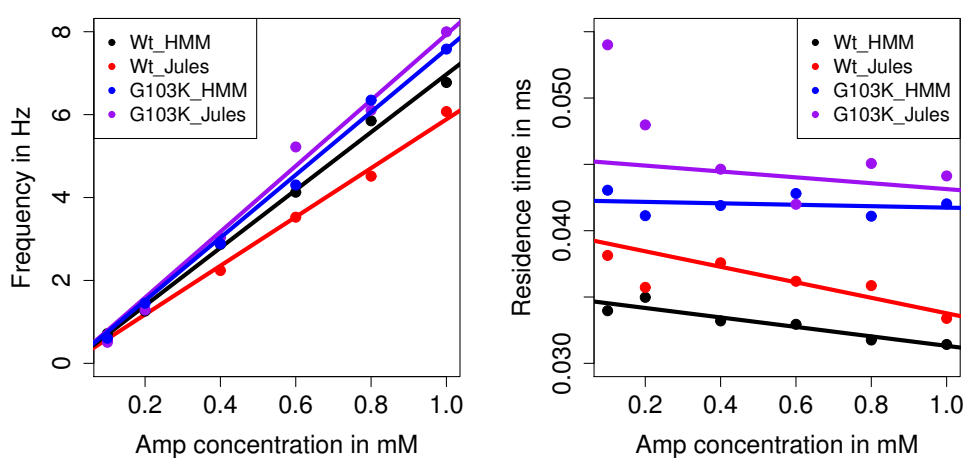

Figure 6.12: Residence times and blockage frequencies at increasing ampicillin concentrations for PorB wild type and PorB G103K. The recordings were performed at 80 mV. For both proteins, the frequencies increase linearly in the ampicillin concentration.

We summarize our findings, a short interpretation is given below, for more details we refer to Bartsch et al. (2017):

- The blockage frequencies increase linearly with the ampicillin concentration.

- The residence time do not dependent significantly on the concentration level of ampicillin.

- The residence times of the mutant are statistically significant larger than the residence times of the wild type. We confirmed this statement by the two-sample Wilcoxon signed-rank test at error level 0.05.

- The blockage frequencies depend linearly on the voltage. However, while for the mutant the frequency is increasing, it is decreasing for the wild type.

- The residence times show a parabolic dependency on the voltage.

Highly simplified, the ampicillin molecules diffuse through the solution and enter the pore if they are close to it and have the necessary orientation. If the number molecules in the solution increases, the time until a blockage occurs decreases. This totally agrees with the linear increase
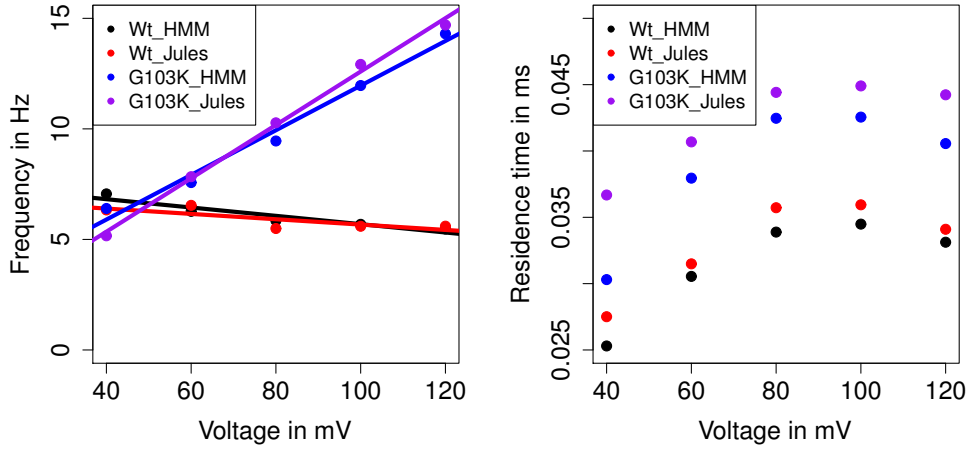
Figure 6.13: Voltage-dependent residence times and blockage frequencies of ampicillin for PorB wild type and PorB G103K in the presence of 1 mM ampicillin. Four measurements were averaged for each protein. For both proteins, the frequencies increase linearly in the applied voltage.

of the blockage frequencies with the ampicillin concentration.

We found no significant dependency of the residence times on the concentration. This confirms the conjecture that a higher concentration of ampicillin molecules in the solution does not effect the single molecule in the pore.

Molecular dynamics (MD) simulations revealed that during the passage through the pore an ampicillin molecule binds in the constriction zone to the channel protein. The binding is similar for the wild-type and the mutant, but the mutant G103K has one additional contact for ampicillin on the extracellular side of the constriction zone, see Figure 9 in Bartsch et al. (2017). This serves as an explanation for the longer residence times of G103K we found as well.

The differences of the porins concerning the dependency between blockage frequency and applied voltage could be caused by multiple reasons. One explanation is that changes in the voltage leads to changes the orientation of the ampicillin molecule in a more favorable or unfavorable way. We refer to Bartsch et al. (2017) for more details.

In general, the membrane of G103K seems to be more resistant concerning the passage of ampicillin molecules, which can explain an antibiotic resistance for cells with the G103K mutation.

6.3.2   Ion channel recordings with varying voltage

The model for ion channel recordings with varying voltage is very similar to the model we developed in Section 6.3.1. The only difference that instead of (6.2), we assume that

$$W_i = (k * (s + (\sigma + \beta_i)\nu)))(t_i) = \int_{-\infty}^{\infty} k(t_i - u)(Y_u + \sigma\nu_u)\mathrm{d}u, \qquad (6.3)$$

where $\beta_i \to 0$ as $n \to \infty$. We stress that we implicitly assumed that the transition probabilities between the states does not depend on the applied voltage, which is doubtful with respect to the findings of Section 6.3.1 concerning the mutant G103K. For the wild type one measurement 67 traces with 1 mM ampicillin concentration and a voltage ramp from 30 mV to 110 mV were recorded. Each trace has $5 \cdot 10^5$ observations. We estimated a blockage frequency of 4.89 Hz and a residence time of 0.033 ms. This totally agrees with the results from Section 6.3.1.

Conclusion and outlook

7.1 Conclusion

Motivated by ion channel recordings with varying voltage we introduce an extended hidden Markov model in this thesis. This trivariate stochastic process $(X_n, Y_n, Z_n)_{n \in \mathbb{N}}$ is characterized by a non-observed homogeneous hidden Markov model $(X_n, Y_n)_{n \in \mathbb{N}}$ and an observed process $(Z_n)_{n \in \mathbb{N}}$. The observation process is inhomogeneous, but the distribution of Z_n is getting "closer" to the distribution of Y_n for increasing n. In Section 2 we give a precise definition and interpretation for this property. We introduce a quasi-maximum likelihood estimator $\theta_{v,n}^{\text{QML}}$ which can be computed without any knowledge about the inhomogenity of the observation process.

The major contribution of this work is Theorem 2.6 and Theorem 2.12 in Section 2. Theorem 2.6 establish the strong consistency of $\theta_{v,n}^{\text{QML}}$, whereas Theorem 2.12 concern the asymptotic normality of $\theta_{v,n}^{\text{QML}}$. Additionally, we prove the same asymptotic results for the maximum likelihood estimator $\theta_{v,n}^{\text{ML}}$ in Corollary 2.7 and Proposition 2.11.

The proof of the asymptotic theory involves a combination of results about maximum likelihood estimation in homogeneous HMMs with results about asymptotic mean stationary processes. In particular, we show that the observed sequence is asymptotically mean stationary (see Theorem 4.4). This results enables us to use the Birkhoff ergodic theorem for $(Z_n)_{n \in \mathbb{N}}$.

Further, we used the Baum-Welch algorithm to compute the quasi-maximum likelihood estimator and showed in a simulation study that this algorithm reaches the asymptotic boundaries of the maximum likelihood estimator. Additionally we developed a forward algorithm for estimation and idealization in filtered data and applied this algorithm to ion channel recordings. We showed a significant difference in the resistance time of ampicillin blockage between the wild type porin PorB and its mutant G103K. These results improve the understanding of potential sources for bacterial resistance and might help to develop new drugs against it to alleviate the severe consequences of multidrug-resistant bacteria.

7.2   Outlook

7.2.1   Model extensions

In our consideration the state space of the underlying Markov chain is finite. Possibly extensions to general state space are of particular interest. For example let for $n, r, l \in \mathbb{N}$ and $S = \mathbb{R}^r$, $G = \mathbb{R}^l$ the dynamics of an extended HMM be given by

$$X_{n+1} = A_\theta X_n + R_\theta U_n,$$

$$Y_n = B_\theta X_n + S_\theta V_n,$$

$$Z_n = Y_n + \beta_n \varepsilon_n,$$

where $(U_n, V_n, \varepsilon_n)$ is an iid sequence of Gaussian vectors with zero mean and identity covariance matrix. We assume that the matrices and random vectors have the appropriate dimensions. If the matrices $A_\theta, B_\theta, R_\theta$ and S_θ satisfy further rank conditions, Douc et al. (2011) proved the strong consistency of the homogeneous model.

Another way to extend our hidden Markov model is to allow time-dependent changes of transition matrix of the underlying Markov chain as in Jensen (2011a) . This might be of particular interest in analyzing ion channel data.

Recall that ion channel recordings are filtered by a Bessel filter. This implies that given the underlying sequence of states, the observations are not independent. Therefore, extensions of our results to so called autoregressive models with Markov regime are mandatory for the analysis of ion channel recordings. Note that Douc et al. (2004) proved consistency and asymptotic normality of the MLE in the homogeneous case.

7.2.2   Condition (2.16)

Recall the Gaussian linear model from Section 3.2 and condition (2.16) from Assumption (CLT1):

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \mathbb{E}_{\theta^*}^\pi \left[S_{n,QML}^{(r)}(\theta^*) \right] = 0, \quad r = 1, \ldots, d. \tag{7.1}$$

In Section 6 we showed that for independent normally distributed random variables this condition is verified, if

$$\beta_n^2 = O(n^{-p}),$$

for some $p > 1/2$. However, this condition is difficult to verify when $K > 1$. In the following we outline a strategy for simplifying (2.16). For $n \in \mathbb{N}$, $v \in \mathcal{P}(S)$ and $\theta \in \Theta$ let $p_{\theta|n}^v$ denote the prediction filter of X_n given the observation $Z_1, \ldots, Z_n$, i.e.,

$$p_{\theta|n}^v(s) = \mathbb{P}_\theta^v(X_n = s \mid Z_1, \ldots, Z_n), \quad s \in S.$$

Further, let $P^v_{\theta|n}$ be the row vector of the prediction filter values, i.e., $P^v_{\theta|n} = (p^v_{\theta|n}(1), \ldots, p^v_{\theta|n}(K))$ and let $(F_{\theta,n})_{n \in \mathbb{N}}$ be a sequence of diagonal matrix with $F_{\theta,n} \in \mathbb{R}^{K \times K}_+$ for all $n \in \mathbb{N}$. The diagonal of $F_{\theta,n}$ is given by

$$\text{diag}(F_{\theta,n}) = (f_{\theta,n}(1, Z_n), \ldots, f_{\theta,n}(K, Z_n))^T.$$

Then for any $n \in \mathbb{N}$ we have

$$P^v_{\theta|n} = \frac{v^T \prod\limits_{i=1}^{n} P_\theta F_{\theta,i}}{\sum\limits_{j=1}^{K} \left(v^T \prod\limits_{i=1}^{n} P_\theta F_{\theta,i} \right)^{(j)}}.$$

Similarly for $n \in \mathbb{N}$ we define

$$Q^v_{\theta|n} = \frac{v^T \prod\limits_{i=1}^{n} P_\theta \tilde{F}_{\theta,i}}{\sum\limits_{j=1}^{K} \left(v^T \prod\limits_{i=1}^{n} P_\theta \tilde{F}_{\theta,i} \right)^{(j)}},$$

where $\tilde{F}_{\theta,n}$ a K-dimensional diagonal matrix with diagonal

$$\text{diag}(\tilde{F}_{\theta,n}) = (f_\theta(1, Z_n), \ldots, f_\theta(K, Z_n)).$$

Suppose now that

$$\max_{j \in S} \left| \frac{(Q^v_{\theta^*|n})^{(j)}}{(P^v_{\theta^*|n})^{(j)}} - 1 \right| = O_{\mathbb{P}^\pi_{\theta^*}}(\alpha_n) \tag{7.2}$$

for some real-valued sequence $(\alpha_n)_{n \in \mathbb{N}}$ with $\alpha_n \to 0$ as $n \to \infty$. Additionally, assume that for any $i, n \in \mathbb{N}$ and $r \in \{1, \ldots, d\}$ we have that

$$\psi^{(r)}(\theta^*, X_{i-1}, X_i, Z_i) \frac{q^{v,1:n}_{\theta^*,(i-1):i}(X_{i-1}, X_i \mid Z_1, \ldots, Z_n)}{p^{v,1:n}_{\theta^*,(i-1):i}(X_{i-1}, X_i \mid Z_1, \ldots, Z_n)}$$

is uniformly integrable w.r.t. $\mathbb{P}^\pi_{\theta^*}$. Then, one can show that

$$E^\pi_{\theta^*} \left[\psi^{(r)}(\theta^*, X_{i-1}, X_i, Z_i) \frac{q^v_{\theta^*,i}(X_{i-1}, X_i \mid Z_1, \ldots, Z_n)}{p^v_{\theta^*,i}(X_{i-1}, X_i \mid Z_1, \ldots, Z_n)} \right] = O(\max(\beta_n^2, \alpha_n))$$

and condition 2.16 is satisfied if

$$\max(\beta_n^2, \alpha_n) = O(n^{-p})$$

for some $p > 1/2$.

Technical proofs

First, we prove a result that specify the "closeness" of Y and Z.

Lemma A.1. *Under the assumption formulated in (C1), we have*

$$\mathbb{P}_\theta^\nu \left(\lim_{n \to \infty} m(Z_n, Y_n) = 0 \right) = 1. \tag{A.1}$$

for any $\theta \in \Theta$ and $\nu \in \mathcal{P}(S)$.

Proof. By (C1) we obtain for any $\varepsilon > 0$ that

$$\sum_{n=1}^\infty \mathbb{P}_\theta^\nu \left(m(Z_n, Y_n) \geq \varepsilon \right) = \sum_{n=1}^\infty \sum_{k=1}^K \mathbb{P}_\theta^\nu \left(m(Z_n, Y_n) \geq \varepsilon, X_n = k \right)$$

$$= \sum_{n=1}^\infty \sum_{k=1}^K \mathbb{P}_\theta^\nu \left(X_n = k \right) \mathbb{P}_\theta \left(m(Z_n, Y_n) \geq \varepsilon \mid X_n = k \right)$$

$$\leq \sum_{n=1}^\infty \max_{k \in S} \mathbb{P}_\theta \left(m(Z_n, Y_n) \geq \varepsilon \mid X_n = k \right) < \infty$$

By the Borel-Cantelli lemma we obtain the desired almost sure convergence of $m(Z_n, Y_n)$ to zero. $\square$

Proof of Theorem 4.4. An intersection-stable generating system of the σ-algebra $\mathcal{B}$ is the union over any finite index set $J \subset \mathbb{N}$ of cylindrical set systems

$$\mathcal{Z}_J := \left\{ \rho_J^{-1}(A_1 \times \cdots \times A_{|J|}) \mid A_j \in \mathcal{B}(G) \text{ open} \right\},$$

where $\rho_J \colon G^{\mathbb{N}} \to G^{|J|}$ is the canonical projection to J, that is, $\rho_J((a_i)_{i \in \mathbb{N}}) = (a_j)_{j \in J}$. By the uniqueness theorem of finite measures it is sufficient to prove for an arbitrary finite index set $J \subset \mathbb{N}$ that for any $B \in \mathcal{Z}_J$ we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{P}_{\theta^*}^{\pi, Z}(T^{-i}(B)) = \mathbb{P}_{\theta^*}^{\pi, Y}(B). \tag{A.2}$$

Fix a finite index set $J = \{j_1, \ldots, j_k\} \subset \mathbb{N}$ and note that $(G^{|J|}, m_J)$, with the metric

$$m_J(a, b) = \sum_{j=1}^{|J|} m(a_j, b_j), \quad a = (a_1, \ldots, a_{|J|}), \ b = (b_1, \ldots, b_{|J|}) \in G^{|J|},$$

is a metric space. Here it is worth to mention that the σ-algebra $\bigotimes_{j \in J} \mathcal{B}(G)$ coincides with the σ-algebra generated by the open sets w.r.t. m_J. By Lemma A.1 we obtain

$$\mathbb{P}_{\theta^*}^\pi \left(\lim_{i \to \infty} m_J \left((Y_{i+j_1}, \ldots, Y_{i+j_k}), (Z_{i+j_1}, \ldots, Z_{i+j_k}) \right) = 0 \right) = 1. \tag{A.3}$$

Let $h : G^{|J|} \to \mathbb{R}$ be a bounded, uniformly continuous function, i.e., for any $\varepsilon > 0$ there is $\delta > 0$ such that for all $a, b \in G^{|J|}$ with $m_J(a, b) < \delta$ we have $|h(a) - h(b)| < \varepsilon$. Then, by the stationarity of Y, the boundedness of h and Fatou's lemma, we have

$$\begin{aligned}
0 &\leq \liminf_{i \to \infty} \mathbb{E}_{\theta^*}^\pi \left[\left| h(Z_{i+j_1}, \ldots, Z_{i+j_k}) - h(Y_{j_1}, \ldots, Y_{j_k}) \right| \right] \\
&\leq \limsup_{i \to \infty} \mathbb{E}_{\theta^*}^\pi \left[\left| h(Z_{i+j_1}, \ldots, Z_{i+j_k}) - h(Y_{i+j_1}, \ldots, Y_{i+j_k}) \right| \right] \\
&\leq \mathbb{E}_{\theta^*}^\pi \left[\limsup_{i \to \infty} \left| h(Z_{i+j_1}, \ldots, Z_{i+j_k}) - h(Y_{i+j_1}, \ldots, Y_{i+j_k}) \right| \right].
\end{aligned} \tag{A.4}$$

By the uniform continuity of h we obtain

$$\lim_{i \to \infty} \left| h(z_{i+j_1}, \ldots, z_{i+j_k}) - h(y_{i+j_1}, \ldots, y_{i+j_k}) \right| = 0$$

for all sequences $((z_{i+j_1}, \ldots, z_{i+j_k}))_{i \in \mathbb{N}}, ((y_{i+j_1}, \ldots, y_{i+j_k}))_{i \in \mathbb{N}} \subset G^{|J|}$ which satisfy

$$\lim_{i \to \infty} m_J((z_{i+j_1}, \ldots, z_{i+j_k}), (y_{i+j_1}, \ldots, y_{i+j_k})) = 0.$$

Then, by using (A.3) we obtain

$$\mathbb{E}_{\theta^*}^\pi \left[\limsup_{i \to \infty} \left| h(Z_{i+j_1}, \ldots, Z_{i+j_k}) - h(Y_{i+j_1}, \ldots, Y_{i+j_k}) \right| \right] \leq 0,$$

such that (by (A.4)) we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\theta^*}^\pi \left[h(Z_{i+j_1}, \ldots, Z_{i+j_k}) \right] = \mathbb{E}_{\theta^*}^\pi \left[h(Y_{j_1}, \ldots, Y_{j_k}) \right].$$

Finally, by Theorem 1.2 in Billingsley (1999) we have for any $A \in \bigotimes_{j \in J} \mathcal{B}(G)$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{P}_{\theta^*}^\pi \left((Z_{i+j_1}, \ldots, Z_{i+j_k}) \in A \right) = \mathbb{P}_{\theta^*}^\pi \left((Y_{j_1}, \ldots, Y_{j_k}) \in A \right),$$

which implies (A.2) for any $B \in \mathcal{Z}_J$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Apart of the fact that we need the previous result to apply Theorem 3 of Barron (1985) it has also the following two useful consequences.

Corollary A.2. *Assume that condition (C1) is satisfied. Then $\mathbb{P}_{\theta^*}^{\pi,Z}$ is ergodic.*

Proof. From Lemma 1 in Leroux (1992) it follows that $\mathbb{P}_{\theta^*}^{\pi,Y}$ is ergodic. Then, the assertion is implied by Theorem 4.4 and by Lemma 7.13 of Gray (2009), which essentially states that $\mathbb{P}_{\theta^*}^{\pi,Y}$ is ergodic if and only if $\mathbb{P}_{\theta^*}^{\pi,Z}$ is ergodic. $\qquad\square$

Corollary A.3. *Assume that condition (C1) is satisfied and let $k \in \mathbb{N}$. Then, for any $g\colon G^k \to \mathbb{R}$ with $\mathbb{E}_{\theta^*}^{\pi}[|g(Y_1,\dots,Y_k)|] < \infty$ we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} g(Z_{j+1},\dots,Z_{j+k}) = \mathbb{E}_{\theta^*}^{\pi}[g(Y_1,\dots,Y_k)], \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.}$$

Proof. By the a.m.s. property and the ergodicity of $\mathbb{P}_{\theta^*}^{\pi,Z}$ the assertion is implied by Theorem 8.1 in Gray (2009). $\qquad\square$

For $z = (z_i)_{i\in\mathbb{N}} \in G^{\mathbb{N}}$ and $k, m \in \mathbb{N}$ with $k < m$ we use $z_{k:m}$ to denote a segment of z. Specifically, let $z_{k:m} = (z_k,\dots,z_m)$.

Let $\lambda_k = \bigotimes_{i=1}^{k} \lambda$ be the product measure of λ with itself, i.e., the measurable space $(G^k, \bigotimes_{i=1}^{k} \mathcal{B}(G))$ is equipped with reference measure λ_k. Now define

$$p_{\theta^*}^{\pi}(z_{1:k} \mid z_{k+1:m}) := \frac{p_{\theta^*}^{\pi}(z_{1:m})}{\int_{G^k} p_{\theta^*}^{\pi}(z_{1:m})\lambda_k(\mathrm{d}z_{1:k})}.$$

We aim to apply Theorem 3 of Barron (1985). For this we need the concept of conditional mutual information.

Definition A.4. *For $k, m, n \in \mathbb{N}$ define the (k, m, n)-conditional mutual information of Z by*

$$I_{k,m}^{Z}(n) := \mathbb{E}_{\theta^*}^{\pi}\left[\log\left(\frac{p_{\theta^*}^{\pi}(Z_{1:k} \mid Z_{k+1:k+m+n})}{p_{\theta^*}^{\pi}(Z_{1:k} \mid Z_{k+1:k+m})}\right)\right].$$

Remark A.5. *Observe that the (k, m, n)-conditional mutual information of Z coincides with the definition of the conditional mutual information of $Z_{k+m+1:k+m+n}$ and $Z_{1:k}$ given $Z_{k+1:k+m}$ in page 1296 of Barron (1985). Note that by Lemma 3 of Barron (1985) it is known that $I_{k,m}^{Z} := \lim_{n\to\infty} I_{k,m}^{Z}(n)$ exists.*

Lemma A.6. *Assume that condition (H4) is satisfied. Then, for every $k, m \in \mathbb{N}$ we have $I_{k,m}^{Z} := \lim_{n\to\infty} I_{k,m}^{Z}(n) < \infty$.*

Proof. For $n \in \mathbb{N}$ we obtain

$$I_{k,m}^{Z}(n) \leq \mathbb{E}_{\theta^*}^{\pi}\left[\left|\log p_{\theta^*}^{\pi}(Z_{1:k} \mid Z_{k+1:k+m})\right|\right]$$
$$+ \mathbb{E}_{\theta^*}^{\pi}\left[\left|\log p_{\theta^*}^{\pi}(Z_{1:k} \mid Z_{k+1:k+m+n})\right|\right].$$

For $1 \le k < j$ we have by inserting $\int_{G^k} \prod_{i=1}^{k} f_{\theta^*,i}(s_i, z_i) \lambda_k(\mathrm{d}z_{1:k}) = 1$ that

$$
\begin{aligned}
p_{\theta^*}^\pi(Z_{1:j}) &= \sum_{s_1,\ldots,s_k \in S} \pi(s_1) \prod_{i=1}^{k} f_{\theta^*,i}(s_i, Z_i) \prod_{i=1}^{k-1} P_{\theta^*}(s_i, s_{i+1}) \\
&\quad \times \sum_{s_{k+1},\ldots,s_{j+1} \in S} P_{\theta^*}(s_k, s_{k+1}) \prod_{\ell=k+1}^{j} f_{\theta^*,\ell}(s_\ell, Z_\ell) P_{\theta^*}(s_\ell, s_{\ell+1}) \\
&\le \max_{x_1,\ldots,x_k \in S} \prod_{i=1}^{k} f_{\theta^*,i}(x_i, Z_i) \int_{G^k} p_{\theta^*}^\pi(z_{1:k}, Z_{k+1:j}) \lambda_k(\mathrm{d}z_{1:k}).
\end{aligned}
$$

By (H4) this leads to

$$
\mathbb{E}_{\theta^*}^\pi \left[\left| \log p_{\theta^*,k|j}^\pi(Z_{1:k} \mid Z_{k+1:j}) \right| \right] \le \max_{x_1,\ldots,x_k \in S} \sum_{i=1}^{k} \mathbb{E}_{\theta^*}^\pi \left[\left| \log \left(f_{\theta^*,i}(x_i, Z_i) \right) \right| \right] < \infty,
$$

which gives $I_{k,m}^Z(n) < \infty$ for any $n \in \mathbb{N}$ and implies the assertion. $\qquad\square$

Theorem A.7. *Assume that the conditions (P1), (H1), (H4) and (C1) are satisfied. Then*

$$
\lim_{n \to \infty} n^{-1} \log p_{\theta^*}^\pi(Z_1, \ldots, Z_n) = \ell(\theta^*) \quad \mathbb{P}_{\theta^*}^\pi\text{-a.s.}
$$

(Recall that $\ell(\theta^*)$ is given by (4.4).)

Proof. Theorem 4.4 shows that $\mathbb{P}_{\theta^*}^{\pi,Z}$ is a.m.s. with stationary mean $\mathbb{P}_{\theta^*}^{\pi,Y}$. Theorem 4.1 yields

$$
\lim_{n \to \infty} n^{-1} \log p_{\theta^*}^\pi(Y_1, \ldots, Y_n) = \ell(\theta^*) \quad \mathbb{P}_{\theta^*}^\pi\text{-a.s.}
$$

Lemma A.6 guarantees that $I_{k,m}^Z < \infty$ for all $k, m \in \mathbb{N}$. Then, the statement follows by Theorem 3 of Barron (1985). $\qquad\square$

The following lemma ensures that the ratio of $p_{\theta^*}^\nu(z_1, \ldots, z_n)$ and $q_{\theta^*}^\nu(z_1, \ldots, z_n)$ does not diverge exponentially or faster.

Lemma A.8. *Assume that condition (C2) is satisfied. Then*

$$
\limsup_{n \to \infty} n^{-1} \log \left(\mathbb{E}_{\theta^*}^\pi \left[\prod_{i=k}^{n} \max_{s \in S} \frac{f_{\theta^*,i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} \right] \right) < 0,
$$

where k is as in assumption (C2).

Proof. The proof is straightforward and follows from the following estimation:

$$\limsup_{n \to \infty} n^{-1} \log \left(\mathbb{E}_{\theta^*}^{\pi} \left[\prod_{i=k}^{n} \max_{s \in S} \frac{f_{\theta^*, i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} \right] \right)$$

$$= \limsup_{n \to \infty} n^{-1} \log \left(\mathbb{E}_{\theta^*}^{\pi} \left[\mathbb{E}_{\theta^*}^{\pi} \left[\prod_{i=k}^{n} \max_{s \in S} \frac{f_{\theta^*, i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} \mid X_k, \dots, X_n \right] \right] \right)$$

$$= \limsup_{n \to \infty} n^{-1} \log \left(\mathbb{E}_{\theta^*}^{\pi} \left[\prod_{i=k}^{n} \mathbb{E}_{\theta^*}^{\pi} \left[\max_{s \in S} \frac{f_{\theta^*, i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} \mid X_i \right] \right] \right)$$

$$\leq \limsup_{n \to \infty} n^{-1} \log \left(\mathbb{E}_{\theta^*}^{\pi} \left[\prod_{i=k}^{n} \max_{s' \in S} \mathbb{E}_{\theta^*}^{\pi} \left[\max_{s \in S} \frac{f_{\theta^*, i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} \mid X_i = s' \right] \right] \right)$$

$$= \limsup_{n \to \infty} n^{-1} \sum_{i=k}^{n} \max_{s' \in S} \log \left(\mathbb{E}_{\theta^*}^{\pi} \left[\max_{s \in S} \frac{f_{\theta^*, i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} \mid X_i = s' \right] \right) \leq 0,$$

where the last line follows from assumption (C2), especially (2.12). $\square$

Corollary A.9. *Assume that condition (C3) is satisfied. Then*

$$\lim_{n \to \infty} n^{-1} \log \left(\mathbb{E}_{\theta^*}^{\pi} \left[\prod_{i=k}^{n} \sup_{\theta' \in \mathcal{E}_\theta} \max_{s \in S} \frac{f_{\theta^*, i}(s, Z_i)}{f_{\theta^*}(s, Z_i)} \right] \right) = 0,$$

where k and $\mathcal{E}_\theta$ are as in (C3).

Proof of Theorem *4.5.* From Theorem A.7 it follows that

$$\lim_{n \to \infty} n^{-1} \log p_{\theta^*}^{\pi}(Z_1, \dots, Z_n) = \ell(\theta^*) \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.} \tag{A.5}$$

and by using (C2) we first show

$$\lim_{n \to \infty} n^{-1} \log q_{\theta^*}^{\pi}(Z_1, \dots, Z_n) = \ell(\theta^*) \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.} \tag{A.6}$$

For any $\varepsilon > 0$ we obtain by Markov's inequality that

$$\mathbb{P}_{\theta^*}^{\pi} \left(n^{-1} \log \left(\frac{q_{\theta^*}^{\pi}(Z_1, \dots, Z_n)}{p_{\theta^*}^{\pi}(Z_1, \dots, Z_n)} \right) \geq \varepsilon \right) = \mathbb{P}_{\theta^*}^{\pi} \left(\frac{q_{\theta^*}^{\pi}(Z_1, \dots, Z_n)}{p_{\theta^*}^{\pi}(Z_1, \dots, Z_n)} \geq \exp(n\varepsilon) \right)$$

$$\leq \exp(-n\varepsilon) \cdot \mathbb{E}_{\theta^*}^{\pi} \left[\frac{q_{\theta^*}^{\pi}(Z_1, \dots, Z_n)}{p_{\theta^*}^{\pi}(Z_1, \dots, Z_n)} \right].$$

By the fact that $\mathbb{E}_{\theta^*}^{\pi} \left[\frac{q_{\theta^*}^{\pi}(Z_1, \dots, Z_n)}{p_{\theta^*}^{\pi}(Z_1, \dots, Z_n)} \right] = 1$, the Borel-Cantelli Lemma implies

$$\limsup_{n \to \infty} n^{-1} \log \left(\frac{q_{\theta^*}^{\pi}(Z_1, \dots, Z_n)}{p_{\theta^*}^{\pi}(Z_1, \dots, Z_n)} \right) \leq 0 \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.}$$

This leads by (A.5) to

$$\limsup_{n\to\infty} n^{-1} \log q_{\theta^*}^{\pi}(Z_1,\ldots,Z_n) \le \ell(\theta^*) \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.} \tag{A.7}$$

Observe that

$$\frac{p_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)}{q_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)} \le \prod_{i=1}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}.$$

Then, with the $k \in \mathbb{N}$ from (C2), in particular (2.10), it follows that

$$\limsup_{n\to\infty} n^{-1} \log\left(\frac{p_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)}{q_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)}\right) \le \limsup_{n\to\infty} n^{-1} \log\left(\prod_{i=1}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right)$$

$$= \limsup_{n\to\infty} n^{-1} \left(\log\left(\prod_{i=1}^{k-1} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right) + \log\left(\prod_{i=k}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right)\right)$$

$$= \limsup_{n\to\infty} n^{-1} \log\left(\prod_{i=k}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right) \quad \mathbb{P}_{\theta}^{\pi}\text{-a.s.}$$

Again, for any $\varepsilon > 0$ we obtain by Markov's inequality that

$$\mathbb{P}_{\theta^*}^{\pi}\left(n^{-1} \log\left(\prod_{i=k}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right) \ge \varepsilon\right)$$

$$= \mathbb{P}_{\theta^*}^{\pi}\left(\prod_{i=k}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)} \ge \exp(n\varepsilon)\right) \le \frac{\mathbb{E}_{\theta^*}^{\pi}\left[\prod_{i=k}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right]}{\exp(n\varepsilon)}$$

$$= \exp\left(n\left(n^{-1} \log\left(\mathbb{E}_{\theta^*}^{\pi}\left[\prod_{i=k}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right]\right) - \varepsilon\right)\right).$$

By Lemma A.8, the Borel-Cantelli Lemma yields

$$\limsup_{n\to\infty} n^{-1} \log\left(\prod_{i=k}^{n} \max_{s\in S} \frac{f_{\theta^*,i}(s,Z_i)}{f_{\theta^*}(s,Z_i)}\right) \le 0 \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.}$$

which leads to

$$\limsup_{n\to\infty} n^{-1} \log\left(\frac{p_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)}{q_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)}\right) \le 0 \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.}$$

This implies

$$\liminf_{n\to\infty} n^{-1} \log\left(\frac{q_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)}{p_{\theta^*}^{\pi}(Z_1,\ldots,Z_n)}\right) \ge 0 \quad \mathbb{P}_{\theta^*}^{\pi}\text{-a.s.} \tag{A.8}$$

By (A.7) and (A.8) we obtain (A.6).

Next we prove the statement of the theorem using (A.6). For any $n \in \mathbb{N}$ observe that

$$
\frac{q_{\theta^*}^{\pi}(Z_1, \ldots, Z_n)}{q_{\theta^*}^{\nu}(Z_1, \ldots, Z_n)} = \frac{\displaystyle\sum_{s_1, \ldots, s_{n+1} \in S} \nu(s_1) \frac{\pi(s_1)}{\nu(s_1)} \prod_{i=1}^{n} f_{\theta^*}(s_i, Z_i) P_{\theta^*}(s_i, s_{i+1})}{\displaystyle\sum_{s_1, \ldots, s_{n+1} \in S} \nu(s_1) \prod_{i=1}^{n} f_{\theta^*}(s_i, Z_i) P_{\theta^*}(s_i, s_{i+1})}
\tag{A.9}
$$
$$
\leq \max_{s \in S} \frac{\pi(s)}{\nu(s)} < \infty,
$$

where the finiteness follows by the fact that ν is strictly positive if and only if π is strictly positive. By using (A.9) we also obtain

$$
\frac{q_{\theta^*}^{\pi}(Z_1, \ldots, Z_n)}{q_{\theta^*}^{\nu}(Z_1, \ldots, Z_n)} \geq \min_{s \in S} \frac{\pi(s)}{\nu(s)} > 0.
\tag{A.10}
$$

Then

$$
\limsup_{n \to \infty} n^{-1} \log q_{\theta^*}^{\nu}(Z_1, \ldots, Z_n)
$$
$$
= \limsup_{n \to \infty} n^{-1} \left(\log \left(\frac{q_{\theta^*}^{\nu}(Z_1, \ldots, Z_n)}{q_{\theta^*}^{\pi}(Z_1, \ldots, Z_n)} \right) + \log q_{\theta^*}^{\pi}(Z_1, \ldots, Z_n) \right)
$$
$$
\leq \limsup_{n \to \infty} n^{-1} \left(\max_{s \in S} \frac{\pi(s)}{\nu(s)} + \log q_{\theta^*}^{\pi}(Z_1, \ldots, Z_n) \right) = \ell(\theta^*)
$$

and by (A.10) we similarly have

$$
\liminf_{n \to \infty} n^{-1} \log q_{\theta^*}^{\nu}(Z_1, \ldots, Z_n) \geq \ell(\theta^*).
$$

By the previous two inequalities the assertion follows.

$\square$

Proof of Theorem *4.7.* By the standard approach to prove consistency, see Lemma B.2 and Theorem B.1, Theorem 4.5 and the fact that

$$
q_{\hat\theta_n}^{\nu}(Z_1, \ldots, Z_n) \geq q_{\theta^*}^{\nu}(Z_1, \ldots, Z_n) \qquad \forall n \in \mathbb{N}
$$

it is sufficient to prove for any closed set $C \subseteq \Theta$ with $\theta^* \notin C$ that

$$
\limsup_{n \to \infty} \sup_{\theta' \in C} n^{-1} \log q_{\theta'}^{\nu}(Z_1, \ldots, Z_n) < \ell(\theta^*) \quad \mathbb{P}_{\theta}^{\pi}\text{-a.s.}
$$

Note that, with η_θ defined in Lemma 4.6, the set $\{B(\theta, \eta_\theta), \theta \in C\}$ is a cover of C. As Θ is compact, C is also compact and thus admits a finite subcover $\{B(\theta_i, \eta_{\theta_i}), \theta_i \in C, i = 1, \ldots, N\}$. Hence it is enough to verify

$$
\limsup_{n \to \infty} \sup_{\theta' \in B(\theta, \eta_\theta) \cap C} n^{-1} \log q_{\theta'}^{\nu}(Z_1, \ldots, Z_n) < \ell(\theta^*) \quad \mathbb{P}_{\theta}^{\pi}\text{-a.s.}
\tag{A.11}
$$

for any $\theta \neq \theta^*$.

Let us fix $\theta \neq \theta^*$ and let η_θ as well as n_θ as in Lemma 4.6. Observe that for any $\theta' \in \Theta$ and any $1 \leq m \leq n$ we have

$$q_{\theta'}^{\nu}(z_1, \ldots, z_n) \leq q_{\theta'}^{\nu}(z_1, \ldots, z_{m-1}) \, q_{\theta'}^{\delta}(z_m, \ldots, z_n), \tag{A.12}$$

$$q_{\theta'}^{\delta}(z_1, \ldots, z_n) \leq q_{\theta'}^{\delta}(z_1, \ldots, z_{m-1}) q_{\theta'}^{\delta}(z_m, \ldots, z_n), \tag{A.13}$$

and define $g_{\theta',m,n}^*(z_m, \ldots, z_n) := \prod_{i=m}^{n} \max_{s \in S} f_{\theta'}(s, z_i)$ as well as $i(n) := \lfloor n/n_\theta \rfloor$.

By using those definitions, and by (A.12) and (A.13) we obtain for sufficiently large $n \in \mathbb{N}$ that

$$
\begin{aligned}
\ell_{\nu,n}(\theta') &\leq \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \ell_{\nu,r}(\theta') + \log q_{\theta'}^{\delta}(Z_{r+1}, \ldots, Z_n) \\
&\leq \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \log g_{\theta',1,r}^*(Z_1, \ldots, Z_r) \\
&\quad + \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \sum_{k=1}^{i(n)-1} \log q_{\theta'}^{\delta}(Z_{n_\theta(k-1)+r+1}, \ldots, Z_{n_\theta k + r}) \\
&\quad + \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \log g_{\theta',n_\theta(i(n)-1)+r+1,n}^*(Z_{n_\theta(i(n)-1)+r+1}, \ldots, Z_n) \\
&= \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \log g_{\theta',1,r}^*(Z_1, \ldots, Z_r) \\
&\quad + \frac{1}{n_\theta} \sum_{r=1}^{n_\theta(i(n)-1)} \log q_{\theta'}^{\delta}(Z_{r+1}, \ldots, Z_{n_\theta+r}) \\
&\quad + \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \sum_{k=n_\theta(i(n)-1)+r+1}^{n} \sup_{s \in S} \log f_{\theta'}(s, Z_k).
\end{aligned}
$$

Observe that for $1 \leq r \leq n_\theta$ holds $n_\theta(i(n) - 1) + r \geq n - 2n_\theta$. Hence we can further estimate the last average and obtain

$$
\begin{aligned}
\sup_{\theta' \in B(\theta, \eta_\theta) \cap C} \ell_{\nu,n}(\theta') &\leq \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \sup_{\theta' \in B(\theta, \eta_\theta) \cap C} \log g_{\theta',1,r}^*(Z_1, \ldots, Z_r) \\
&\quad + \frac{1}{n_\theta} \sum_{r=1}^{n_\theta(i(n)-1)} \sup_{\theta' \in B(\theta, \eta_\theta) \cap C} \log q_{\theta'}^{\delta}(Z_{r+1}, \ldots, Z_{n_\theta+r}) \\
&\quad + \sum_{k=n-2n_\theta+1}^{n} \sup_{\theta' \in B(\theta, \eta_\theta) \cap C} \max_{s \in S} \log (f_{\theta'}(s, Z_k))^+ .
\end{aligned}
$$

We multiply both sides of the previous inequality by n^{-1} and consider the limit $n \to \infty$ of each sum on the right-hand side. In particular we show that this is smaller than $\ell(\theta^*)$ which verifies (A.11).

To the first sum: By the fact that $\int_G f_{\theta'}(s,z)\lambda(dz) = 1$, for any $s \in S$ we conclude

$$\lambda(\{z \in G : f_\theta(s,z) = \infty\}) = 0.$$

Hence

$$\mathbb{P}_{\theta^*}^\pi (f_{\theta'}(s, Z_i) = \infty) = 0,$$

and (H3) implies

$$\mathbb{P}_{\theta^*}^\pi \left(\sup_{\theta' \in B(\theta,\eta_\theta) \cap C} \log g_{\theta',1,r}^*(Z_1,\ldots,Z_r) = \infty \right) = 0 \qquad \forall r \in \mathbb{N}.$$

This leads to

$$\lim_{n \to \infty} \frac{1}{n} \frac{1}{n_\theta} \sum_{r=1}^{n_\theta} \sup_{\theta' \in B(\theta,\eta_\theta) \cap C} \log g_{\theta',1,r}^*(Z_1,\ldots,Z_r) = 0 \qquad \mathbb{P}_{\theta^*}^\pi\text{-a.s.}$$

To the second sum: By the fact that $i(n)/n \to n_\theta^{-1}$ as $n \to \infty$, Lemma 4.6 and Corollary A.2 we obtain

$$\lim_{n \to \infty} \frac{1}{n} \frac{1}{n_\theta} \sum_{r=1}^{n_\theta(i(n)-1)} \sup_{\theta' \in B(\theta,\eta_\theta) \cap C} \log q_{\theta'}^\delta(Z_{r+1},\ldots,Z_{n_\theta+r})$$
$$= \frac{1}{n_\theta} \mathbb{E}_{\theta^*}^\pi \left[\sup_{\theta' \in B(\theta,\eta_\theta) \cap C} \log p_{\theta'}^\delta(Y_1,\ldots,Y_{n_\theta}) \right] < \ell(\theta^*).$$

To the third sum: By assumption (H2) it follows that

$$\mathbb{E}_{\theta^*}^\pi \left[\sup_{\theta' \in \mathcal{U}_\theta} \max_{s \in S} (\log f_\theta(s, Y_1))^+ \right] \leq \sum_{s \in S} \mathbb{E}_{\theta^*}^\pi \left[\sup_{\theta' \in \mathcal{U}_\theta} (\log f_\theta(s, Y_1))^+ \right] < \infty.$$

and by Corollary A.2 we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^n \sup_{\theta' \in B(\theta,\eta_\theta)} \max_{s \in S} \log(f_{\theta'}(s, Z_k))^+ = \mathbb{E}_{\theta^*}^\pi \left[\sup_{\theta' \in B(\theta,\eta_\theta)} \max_{s \in S} (\log f_\theta(s, Y_1))^+ \right].$$

Hence

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=n-2n_\theta+1}^n \sup_{\theta' \in B(\theta,\eta_\theta)} \max_{s \in S} \log(f_{\theta'}(s, Z_k))^+ = 0 \qquad \mathbb{P}_{\theta^*}^\pi\text{-a.s.}$$

and the proof is complete . $\square$

The following proofs concern Theorem 2.12.

Proof of Corollary *4.14.* Note that

$$
\Big| \sum_{s_{i-1}, s_i \in S} \sum_{s_r, s_t \in S} \Big| q_{\theta,(i-1):i|\{r,l\}}^{\nu,r:l}(s_{i-1}, s_i \mid s_r, s_l, Z_r, \dots Z_l)
$$
$$
\times \big(q_{\theta,\{r,l\}}^{\nu,r:l}(s_r, s_l \mid Z_r, \dots, Z_l) - q_{\theta,\{r:l\}}^{\nu,1:n}(s_r, s_l \mid Z_1, \dots, Z_n) \big) \Big|
$$
$$
= 2 \sup_{B \in \mathcal{S}^2} \Big| \sum_{s_r, s_l \in S} \sum_{s \in B} q_{\theta,(i-1):i|\{r,l\}}^{\nu,r:l}(s \mid s_r, s_l, Z_r, \dots Z_l) q_{\theta,\{r,l\}}^{\nu,r:l}(s_r, s_l \mid Z_r, \dots, Z_l)
$$
$$
- \sum_{s_r, s_l \in S} \sum_{s \in B} q_{\theta,(i-1):i|\{r,l\}}^{\nu,r:l}(s \mid s_r, s_l, Z_r, \dots Z_l) q_{\theta,\{r,l\}}^{\nu,1:n}(s_r, s_l \mid Z_1, \dots, Z_n) \Big|
$$
$$
\leq 2 \sup_{B \in \mathcal{S}^2} \Big(\sup_{s_r, s_l \in S} \sum_{s \in B} q_{\theta,r:l|\{r,l\}}^{\nu,r:l}(s \mid s_r, s_l, Z_r, \dots Z_l) - \inf_{s_r, s_l \in S} \sum_{s \in B} q_{\theta,r:l|\{r,l\}}^{\nu,r:l}(s \mid s_r, s_l, Z_r, \dots Z_l) \Big)
$$
$$
\leq 2 \sup_{B \in \mathcal{S}^2} \Big(\sup_{s_r, s_l \in S} \sum_{s \in B} q_{\theta,(i-1):i|\{r,l\}}^{\nu,r:l}(s \mid s_r, s_l, Z_1, \dots Z_n) - \inf_{s_r, s_l \in S} \sum_{s \in B} q_{\theta,(i-1):i|\{r,l\}}^{\nu,r:l}(s \mid s_r, s_l, Z_1, \dots Z_n) \Big).
$$

The assertion follows from Corollary 4.12.

$\square$

Proof of Lemma *4.15.* Let $\theta \in B(\theta^*, \delta_0)$. Similarly to (4.9) one can show that

$$
\Big| \mathbb{E}_\theta^\pi \Big[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})}{p_{\theta,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})} \mid Z_{1:n} \Big]
$$
$$
- \mathbb{E}_\theta^\pi \Big[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \Big] \Big|
$$
$$
\leq b^0(Z_{r:t}) 4 \rho^k
$$

and

$$
\Big| \mathbb{E}_{\theta^*}^\pi \Big[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta^*,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})}{p_{\theta^*,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})} \mid Z_{1:n} \Big]
$$
$$
- \mathbb{E}_{\theta^*}^\pi \Big[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta^*,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta^*,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \Big] \Big|
$$
$$
\leq b^0(Z_{r:t}) 4 \rho^k
$$

Therefore

$$\left| \mathbb{E}_\theta^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})}{p_{\theta,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})} \mid Z_{1:n} \right] - \mathbb{E}_{\theta^*}^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta^*,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})}{p_{\theta^*,r:t}^{\nu,1:n}(X_{r:t} \mid Z_{1:n})} \mid Z_{1:n} \right] \right|$$

$$\leq \left| \mathbb{E}_\theta^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \right] \right.$$

$$\left. - \mathbb{E}_{\theta^*}^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta^*,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta^*,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \right] \right|$$

$$+ b^0(Z_{r:t}) 8\rho^k$$

To complete the proof, we will show that

$$\left| \mathbb{E}_\theta^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \right] \right.$$

$$\left. - \mathbb{E}_{\theta^*}^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta^*,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta^*,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k} Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \right] \right|$$

$$\leq b^0(Z_{r:t}) 2\rho \left| \theta - \theta^* \right| \sum_{i=r-k}^{t+k} h(Z_i).$$

Note that

$$\mathbb{E}_\theta^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \right]$$

$$= \sum_{s_r,\dots,s_t \in S} b(s_{r:t}, Z_{r:t}) q_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(s_r, \dots, s_t \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)}),$$

which can be used to show that for any $u \in \{1, \dots, d\}$ we have

$$\left| \frac{\partial}{\partial \theta^{(u)}} \mathbb{E}_\theta^\pi \left[b(X_{r:t}, Z_{r:t}) \frac{q_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t\mid\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)} \right] \right|$$

$$\leq 2b^0(Z_r, \dots, Z_t) \sum_{i=r-k}^{t+k} h_i(Z_i).$$

Finally, note that

$$
\mathbb{E}_{\theta}^{\pi}\left[b(X_{r:t}, Z_{r:t})\frac{q_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)}\right]
$$

$$
- \mathbb{E}_{\theta^*}^{\pi}\left[b(X_{r:t}, Z_{r:t})\frac{q_{\theta^*,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta^*,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})} \mid X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)}\right]
$$

$$
= \int_0^1 \frac{\partial}{\partial u}\mathbb{E}_{\theta^*+u(\theta-\theta^*)}^{\pi}\left[b(X_{r:t}, Z_{r:t})\frac{q_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t}|X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t}|X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}|X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)}\right]\mathrm{d}u
$$

$$
= (\theta-\theta^*)\int_0^1 \frac{\partial}{\partial\theta}\mathbb{E}_{\theta}^{\pi}\left[b(X_{r:t}, Z_{r:t})\frac{q_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t}|X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}{p_{\theta,r:t|\{r-k,t+k\}}^{\nu,(r-k):(t+k)}(X_{r:t}|X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)})}|X_{r-k}, X_{t+k}, Z_{(r-k):(t+k)}\right]\mathrm{d}u
$$

<div align="right">□</div>

Proof of Proposition 4.16. Fix an integer $i \in \mathbb{N}$. For $n \in \mathbb{N}$ and $\delta \in B(\theta^*, \delta_n)$ we find that

$$
\left|\mathbb{E}_{\theta}^{\pi}\left[a_i(\theta, X_{(i-1):i}, Z_i)\frac{q_{\theta,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta,(i-1):i}^{\nu,1,:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n}\right]\right.
$$

$$
\left.- \mathbb{E}_{\theta}^{\pi}\left[a_i(\theta^*, X_{(i-1):i}, Z_i)\frac{q_{\theta,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n}\right]\right|
$$

$$
\leq \delta_n \bar{a}_i(Z_i). \tag{A.14}
$$

With Lemma 4.15 it follows that for any $l \in \mathbb{N}$ with $i - l > 0$ and $i + l < n$ we have

$$
\left|\mathbb{E}_{\theta}^{\pi}\left[a_i(\theta^*, X_{(i-1):i}, Z_i)\frac{q_{\theta,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n}\right]\right.
$$

$$
\left.- \mathbb{E}_{\theta^*}^{\pi}\left[a_i(\theta^*, X_{(i-1):i}, Z_i)\frac{q_{\theta,(i-1):i}^{\nu}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta,(i-1):i}^{\nu}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n}\right]\right|
$$

$$
\leq a_i^0(Z_i)\left(2\rho\delta_n \sum_{m=i-l}^{i+l} h(Z_i) + 8\rho^l\right). \tag{A.15}
$$

The estimates (A.14) and (A.15) imply that

$$
\lim_{n\to\infty}\mathbb{E}_{\theta^*}^{\pi}\left[\sup_{\theta\in B(\theta^*,\delta_n)}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}_{\theta}^{\pi}\left[a_i(\theta, X_{(i-1):i}, Z_i)\frac{q_{\theta,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta,(i-1):i}^{\nu}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n}\right]\right.\right.\right.
$$

$$
\left.\left.\left.-\mathbb{E}_{\theta^*}^{\pi}\left[a_i(\theta^*, X_{(i-1):i}, Z_i)\frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n}\right]\right)\right|\right]
$$

$$
\leq \lim_{n\to\infty}\mathbb{E}_{\theta^*}^{\pi}\left[\frac{1}{n}\sum_{i=1}^{n}\delta_n\bar{a}_i(Z_i) + a_i^0(Z_i)\left(2\rho\delta_n \sum_{m=i-l}^{i+l} h(Z_i) + 8\rho^l\right)\right] \tag{A.16}
$$

The Cauchy-Schwartz inequality implies that there exists a constant K such that for all $i \in \mathbb{N}$ we have $\mathbb{E}^\pi_{\theta^*}[\bar{a}_i(Z_i)] < K$, $\mathbb{E}^\pi_{\theta^*}\left[a^0_i(Z_i)\right] < K$ and $\mathbb{E}^\pi_{\theta^*}\left[a^0_i(Z_i)h_u(Z_u)\right] < K$. Finally, we can bound (A.16) by

$$\delta_n K + 2\rho\delta_n 2(l+1)K + 8\rho^l K.$$

Choosing $l = \lfloor \delta_n^{-1/2} \rfloor$ gives the desired result as $n \to \infty$. $\qquad\square$

For $i \in \mathbb{N}$ let $a_i, b_i : \Theta \times S \times S \times G \to \mathbb{R}$ be functions, we write $a_i(\theta) = a_i(\theta, X_{i-1}, X_i, Z_i)$ and $b_i(\theta) = b_i(\theta, X_{i-1}, X_i, Z_i)$ in the following.

Lemma A.10. *For $i \in \mathbb{N}$ let $a_i, b_i : \Theta \times S \times S \times G \to \mathbb{R}$ be functions and let $(a_i)_{i\in\mathbb{N}}$ and $(b_i)_{i\in\mathbb{N}}$ belong to the class $C_{3,1}$. Further, let the functions $(h)_{i\in\mathbb{N}}$ belong to the class C_3, where h is defined in (4.10). Then there exist constants q_2 and q_3 such that for $\delta > 0$ and any integer l, v, u with $u - l \geq 1$ and $v + l \leq n$ and $u \leq v$, we have*

$$\mathbb{E}^\pi_{\theta^*}\left[\sup_{|\theta-\theta^*|\leq\delta}\left|\mathrm{Cov}^v_\theta\left(a_u(\theta)\frac{q^{v,1:n}_{\theta,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}{p^v_{\theta,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta)\frac{q^{v,1:n}_{\theta,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})}{p^{v,1:n}_{\theta,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right)\right.\right.$$
$$\left.\left.-\mathrm{Cov}^v_{\theta^*}\left(a_u(\theta^*)\frac{q^{v,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}{p^{v,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*)\frac{q^{v,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})}{p^{v,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right)\right|\right]$$
$$\leq \delta\left(2q_2 + 2\rho q_3\left[|v-u| + 6(l+1)\right]\right) + 24q_2\rho^l.$$

Proof. Note that for any random variables $W_1, W_2 : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with finite second moment it holds that

$$\mathrm{Cov}(W_1, W_2) \leq (\mathrm{Var}(W_1))^{1/2}(\mathrm{Var}(W_2))^{1/2}.$$

This and the triangular inequality imply that

$$\left|\mathrm{Cov}^v_\theta\left(a_u(\theta)\frac{q^{v,1:n}_{\theta,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}{p^{v,1:n}_{\theta,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta)\frac{q^{v,1:n}_{\theta,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})}{p^{v,1:n}_{\theta,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right)\right.$$
$$\left.-\mathrm{Cov}^v_\theta\left(a_u(\theta^*)\frac{q^{v,1:n}_{\theta,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}{p^{v,1:n}_{\theta,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*)\frac{q^{v,1:n}_{\theta,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})}{p^{v,1:n}_{\theta,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right)\right|$$
$$\leq \delta(\bar{a}_u(Z_u)b^0_v(Z_v) + a^0_u(Z_u)\bar{b}_v(Z_v)).$$

It follows that

$$\mathbb{E}^\pi_{\theta^*}\left[\left|\mathrm{Cov}^v_\theta\left(a_u(\theta)\frac{q^v_{\theta,u}(X_{u-1:u} \mid Z_{1:n})}{p^v_{\theta,u}(X_{u-1:u} \mid Z_{1:n})}, b_v(\theta)\frac{q^v_{\theta,v}(X_{v-1:v} \mid Z_{1:n})}{p^v_{\theta,v}(X_{v-1:v} \mid Z_{1:n})} \mid Z_{1:n}\right)\right.\right.$$
$$\left.\left.-\mathrm{Cov}^v_\theta\left(a_u(\theta^*)\frac{q^v_{\theta,u}(X_{u-1:u} \mid Z_{1:n})}{p^v_{\theta,u}(X_{u-1:u} \mid Z_{1:n})}, b_v(\theta^*)\frac{q^v_{\theta,v}(X_{v-1:v} \mid Z_{1:n})}{p^v_{\theta,v}(X_{v-1:v} \mid Z_{1:n})} \mid Z_{1:n}\right)\right|\right]$$
$$\leq 2\delta q_2,$$

where q_2 is an upper bound on the second moments of $\bar{a}_u(Z_u)$, $b^0_v(Z_v)$, $a^0_u(Z_u)$ and $\bar{b}_v(Z_v)$ for all

$u, v \in \mathbb{N}$. Lemma 4.15 implies that for any $l \in \mathbb{N}$ with $u - l \geq 1$ and $v + l \leq n$ we can bound

$$\left| \mathbb{E}_\theta^\pi \left[a_u(\theta^*) b_v(\theta^*) \frac{q_{\theta,(u-1):v}^{v,1:n}(X_{(u-1):v} \mid Z_{1:n})}{p_{\theta,(u-1):v}^{v,1:n}(X_{(u-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right] \right.$$

$$\left. - \mathbb{E}_{\theta^*}^\pi \left[a_u(\theta^*) b_v(\theta^*) \frac{q_{\theta^*,(u-1):v}^{v,1:n}(X_{u-1:v} \mid Z_{1:n})}{p_{\theta^*,(u-1):v}^{v,1:n}(X_{(u-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right] \right|$$

by

$$a_u^0(Z_u) b_v^0(Z_v) \left(2\rho\delta \sum_{i=u-l}^{v+l} h(Z_i) + 8\rho^l \right). \tag{A.17}$$

Again Lemma 4.15 and the triangular inequality show that

$$\left| \mathbb{E}_\theta^\pi \left[a_u(\theta^*) \frac{q_{\theta,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})} \mid Z_{1:n} \right] \mathbb{E}_\theta^\pi \left[b_v(\theta^*) \frac{q_{\theta,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right] \right.$$

$$\left. - \mathbb{E}_{\theta^*}^\pi \left[a_u(\theta^*) \frac{q_{\theta^*,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})} \mid Z_{1:n} \right] \mathbb{E}_{\theta^*}^\pi \left[b_v(\theta^*) \frac{q_{\theta^*,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right] \right|$$

$$\leq a_u(Z_u)^0 b_v^0(Z_v) \left(2\rho\delta \left(\sum_{i=u-l}^{u+l} h(Z_i) + \sum_{i=v-l}^{v+l} h(Z_i) \right) + 16\rho^l \right) \tag{A.18}$$

for any integer l with $u - l \geq 1$ and $v + l \leq n$. Combining (A.17) and (A.18) we have that

$$\left| \mathrm{Cov}_\theta^v \left(a_u(\theta^*) \frac{q_{\theta,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*) \frac{q_{\theta,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right) \right.$$

$$\left. - \left| \mathrm{Cov}_{\theta^*}^v \left(a_u(\theta^*) \frac{q_{\theta^*,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*) \frac{q_{\theta^*,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right) \right| \right.$$

$$\leq a_u^0(Z_u) b_v^0(Z_v) \left(2\rho\delta \left(\sum_{i=u-l}^{v+l} h_i(Z_i) + \sum_{i=u-l}^{u+l} h_i(Z_i) + \sum_{i=v-l}^{v+l} h_i(Z_i) \right) + 24\rho^l \right). \tag{A.19}$$

Hölder's inequality implies that the mean of (A.19) is bounded by

$$2\rho\delta q_3 \left(|v - l| + 6(l + 1) \right) + 24q_2\rho^l,$$

where q_3 is a bound on the third moments of $h(Z_i), a_u^0(Z_u), b_v^0(Z_v)$ for all $i, u, v \in \mathbb{N}$. The triangular inequality proves the Lemma. □

Proof of Proposition 4.17. Similarly to Theorem 17.2.1 in Ibragimov and Linnik (1971) one can show with Corollary 4.12 that for any $\theta \in B(\theta, \delta_0)$ we have

$$\left| \mathrm{Cov}_\theta^v \left(a_u(\theta) \frac{q_{\theta,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta,(u-1):u}^{v,1:n}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta) \frac{q_{\theta,(v-1):v}^{v,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta,(v-1):v}^{v,1:v}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right) \right| \leq 4a_u^0 b_v^0 \rho^{|v-u|-3}$$

$$\tag{A.20}$$

and therefore

$$\mathbb{E}_{\theta^*}^{\pi}\left[\left|\mathrm{Cov}_{\theta}^{\gamma}\left(a_u(\theta)\frac{q_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})}{p_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})},b_v(\theta)\frac{q_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}{p_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}\mid Z_{1:n}\right)\right|\right]\le 4q_2\rho^{|v-u|-3},$$

where q_2 is defined as in Lemma *A*.10. For $l\in\mathbb{N}$ we get that

$$\mathbb{E}_{\theta^*}^{\pi}\left[\sup_{|\theta-\theta^*|\le\delta_n}\left|\frac{1}{n}\sum_{u,v=1}^{n}\left(\mathrm{Cov}_{\theta}^{\gamma}\left(a_u(\theta)\frac{q_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})}{p_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})},b_v(\theta)\frac{q_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}{p_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}\mid Z_{1:n}\right)\right.\right.\right.$$
$$\left.\left.\left.-\mathrm{Cov}_{\theta^*}^{\gamma}\left(a_u(\theta^*)\frac{q_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})},b_v(\theta^*)\frac{q_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}\mid Z_{1:n}\right)\right)\right|\right]$$

$$\le\frac{1}{n}\sum_{u=1}^{n}\sum_{\substack{v\in\{1,\dots,n\}:\\|v-u|>l}}\mathbb{E}_{\theta^*}^{\pi}\left[\sup_{|\theta-\theta^*|\le\delta_n}\left|\left(\mathrm{Cov}_{\theta}^{\gamma}\left(a_u(\theta)\frac{q_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}|Z_{1:n})}{p_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}|Z_{1:n})},b_v(\theta)\frac{q_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}|Z_{1:n})}{p_{\theta,(v-1):v}^{\gamma}(X_{(v-1):v}|Z_{1:n})}|Z_{1:n}\right)\right.\right.\right.$$
$$\left.\left.\left.-\mathrm{Cov}_{\theta^*}^{\gamma}\left(a_u(\theta^*)\frac{q_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})},b_v(\theta^*)\frac{q_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}\mid Z_{1:n}\right)\right)\right|\right]$$

$$+\sum_{\substack{v\in\{1,\dots,n\}:\\|v-u|\le l}}\mathbb{E}_{\theta^*}^{\pi}\left[\sup_{|\theta-\theta^*|\le\delta_n}\left|\left(\mathrm{Cov}_{\theta}^{\gamma}\left(a_u(\theta)\frac{q_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}|Z_{1:n})}{p_{\theta,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}|Z_{1:n})},b_v(\theta)\frac{q_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}|Z_{1:n})}{p_{\theta,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}|Z_{1:n})}|Z_{1:n}\right)\right.\right.\right.$$
$$\left.\left.\left.-\mathrm{Cov}_{\theta^*}^{\gamma}\left(a_u(\theta^*)\frac{q_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\gamma,1:n}(X_{(u-1):u}\mid Z_{1:n})},b_v(\theta^*)\frac{q_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}|Z_{1:n})}{p_{\theta^*,(v-1):v}^{\gamma,1:n}(X_{(v-1):v}\mid Z_{1:n})}\mid Z_{1:n}\right)\right)\right|\right]$$

$$\le\frac{1}{n}\sum_{u=1}^{n}\sum_{\substack{v\in\{1,\dots,n\}:\\|v-u|>l}}8q_2\rho^{|v-u|-3}$$
$$+\sum_{\substack{v\in\{1,\dots,n\}:\\|v-u|\le l}}\delta_n\left(2q_2+2\rho q_3(|v-u|+6(l+1)+24q_2\rho^l)\right)$$

$$\le\frac{1}{n}\sum_{u=1}^{n}8q_2\frac{\rho^{l-3}}{1-\rho}+(2l+1)\delta_n\left(2q_2+2\rho q_3(|v-u|+6(l+1)+24q_2\rho^l)\right)$$

$$=8q_2\frac{\rho^{l-3}}{1-\rho}+(2l+1)\delta_n\left(2q_2+2\rho q_3(|v-u|+6(l+1)+24q_2\rho^l)\right)$$

Choosing $l=\lfloor\delta_n^{-1/4}\rfloor$ completes the proof. □

Proof of Lemma 4.19. Let $u,l,n\in\mathbb{N}$ with $u-l\ge 1$ and $u+l\le n$. From the argument we used

in Corollary 4.14 we have that

$$
\left| \mathbb{E}^{\pi}_{\theta^*} \left[a_u(\theta^*, X_{(u-1):u}, Z_u) \frac{q^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})} \mid Z_{1:n} \right] \right.
$$
$$
\left. - \mathbb{E}^{\pi}_{\theta^*} \left[a_u(\theta^*, X_{(u-1):u}, Z_u) \frac{q^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{(u-l):(u+l)})}{p^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{(u-l):(u+l)})} \mid Z_{(u-l):(u+l)} \right] \right|
$$
$$
\leq 4 a^0_u(Z_u).
$$

It follows that

$$
\mathrm{Cov}^{\pi}_{\theta^*} \left(\mathbb{E}^{\pi}_{\theta^*} \left[a_u(\theta^*, X_{(u-1):u}, Z_u) \frac{q^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{1:n})} \mid Z_{1:n} \right], \right.
$$
$$
\left. \mathbb{E}^{\pi}_{\theta^*} \left[a_v(\theta^*, X_{(v-1):v}, Z_v) \frac{q^{\nu,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})}{p^{\nu,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n} \right] \right)
$$
$$
= \mathrm{Cov}^{\pi}_{\theta^*} \left(\mathbb{E}^{\pi}_{\theta^*} \left[a_u(\theta^*, X_{(u-1):u}, Z_u) \frac{q^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{(u-l):(u+l)})}{p^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{(u-l):(u+l)})} \mid Z_{(u-l):(u+l)} \right], \right.
$$
$$
\left. E^{\pi}_{\theta^*} \left[a_v(\theta^*, X_{(v-1):v}, Z_v) \frac{q^{\nu,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{(v-l):(v+l)})}{p^{\nu,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{(v-l):(v+l)})} \mid Z_{(v-l):(v+l)} \right] \right) + O(q_2 \rho^l),
$$

where q_2 is again an upper bound for the second moment of $a^0_u(Z_u)$ for all $u \in \mathbb{N}$. Using the mixing of the observed process proven in Lemma 4.13 we can show as in Theorem 17.2.2 in Ibragimov and Linnik (1971) that

$$
\mathrm{Cov}^{\pi}_{\theta^*} \left(\mathbb{E}^{\pi}_{\theta^*} \left[a_u(\theta^*, X_{(u-1):u}, Z_u) \frac{q^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{(u-l):(u+l)})}{p^{\nu,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u} \mid Z_{(u-l):(u+l)})} \mid Z_{(u-l):(u+l)} \right], \right.
$$
$$
\left. E^{\pi}_{\theta^*} \left[a_v(\theta^*, X_{(v-1):v}, Z_v) \frac{q^{\nu,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{(v-l):(v+l)})}{p^{\nu,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v} \mid Z_{(v-l):(v+l)})} \mid Z_{(v-l):(v+l)} \right] \right)
$$
$$
= O(q_3 \rho^{\frac{\max(0,|v-u|-2l)}{3}}),
$$

where q_3 is an upper bound on the third moment of $a^0_u(Z_u)$ for all $u \in \mathbb{N}$. Taking $l = \lfloor (v - u)/4 \rfloor$,

we find that

$$
\begin{aligned}
&\operatorname{Var}_{\theta^*}^{\pi}\left(\sum_{u=1}^{n} \mathbb{E}_{\theta^*}^{\pi}\left[a_i(\theta^*, X_{(i-1):i}, Z_i) \frac{q_{\theta^*,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})}{p_{\theta^*,(i-1):i}^{\nu,1:n}(X_{(i-1):i} \mid Z_{1:n})} \mid Z_{1:n}\right]\right) \\
&= \operatorname{Cov}_{\theta^*}^{\pi}\left(\sum_{u=1}^{n} \mathbb{E}_{\theta^*}^{\pi}\left[a_u(\theta^*, X_{(u-1):u}, Z_u) \frac{q_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})} \mid Z_{1:n}\right],\right. \\
&\qquad\qquad \left.\sum_{v=1}^{n} \mathbb{E}_{\theta^*}^{\pi}\left[a_v(\theta^*, X_{(v-1):v}, Z_v) \frac{q_{\theta^*,(v-1):v}^{\nu,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\nu,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right]\right) \\
&= \sum_{u=1}^{n}\sum_{v=1}^{n} \operatorname{Cov}_{\theta^*}^{\pi}\left(\mathbb{E}_{\theta^*}^{\pi}\left[a_u(\theta^*, X_{(u-1):u}, Z_u) \frac{q_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})} \mid Z_{1:n}\right],\right. \\
&\qquad\qquad \left.\mathbb{E}_{\theta^*}^{\pi}\left[a_v(\theta^*, X_{(v-1):v}, Z_v) \frac{q_{\theta^*,(v-1):v}^{\nu,1:n}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\nu,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right]\right) \\
&= \sum_{u=1}^{n}\sum_{v=1}^{n} O\left(q_3 \rho^{|v-u|/6} + q_2 \rho^{|v-u|/4}\right) = O(n)
\end{aligned}
$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

Proof of Lemma 4.20. The proof is similar to the proof Lemma 4.19. For $i \in \mathbb{N}$ we write a_i for $a_i(\theta^*, X_{i-1}, X_i, Z_i)$. Let $u, v, z, l \in \mathbb{N}$ with $u - l \geq 1$ and $u + l \leq n$. Further let

$$
\xi_u = \sum_{v=1}^{n} \operatorname{Cov}_{\theta^*}^{\nu}\left(a_u(\theta^*) \frac{q_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*) \frac{q_{\theta^*,(v-1):v}^{\nu}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\nu,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right)
$$

and

$$
\xi_u^l = \sum_{v=u-l}^{u+l} \operatorname{Cov}_{\theta^*}^{\nu}\left(a_u(\theta^*) \frac{q_{\theta^*,(u-1):u}^{\nu}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*) \frac{q_{\theta^*,(v-1):v}^{\nu}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\nu,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right).
$$

In the following we use the abbreviation

$$
\begin{aligned}
&\operatorname{Cov}_{\theta^*}^{\nu}(a_u, b_v \mid Z_{1:n}) \\
&= \operatorname{Cov}_{\theta^*}^{\nu,1:n}\left(a_u(\theta^*) \frac{q_{\theta^*,(u-1):u}^{\nu,1:n}(X_{(u-1):u} \mid Z_{1:n})}{p_{\theta^*,(u-1):u}^{\nu}(X_{(u-1):u} \mid Z_{1:n})}, b_v(\theta^*) \frac{q_{\theta^*,(v-1):v}^{\nu}(X_{(v-1):v} \mid Z_{1:n})}{p_{\theta^*,(v-1):v}^{\nu,1:n}(X_{(v-1):v} \mid Z_{1:n})} \mid Z_{1:n}\right).
\end{aligned}
$$

Note that

$$\text{Cov}_{\theta^*}^{\pi}(\xi_u, \xi_z) - \text{Cov}_{\theta^*}^{\pi}\left(\xi_u^l, \xi_z^l\right)$$

$$= \sum_{v=1}^{n} \sum_{w=1}^{n} \text{Cov}_{\theta^*}^{\pi}\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n}), \text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right)$$

$$- \sum_{v=u+-l}^{u+l} \sum_{w=z-l}^{z+l} \text{Cov}_{\theta^*}^{\pi}\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n}), \text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right)$$

$$= \sum_{v=u-l}^{u+l} \sum_{\substack{w\in\{1,\ldots,n\} \\ |w-z|>l}} \text{Cov}_{\theta^*}^{\pi}\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n}), \text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right)$$

$$+ \sum_{\substack{v\in\{1,\ldots,n\} \\ |v-u|>l}} \sum_{w=z-l}^{z+l} \text{Cov}_{\theta^*}^{\pi}\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n}), \text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right)$$

$$+ \sum_{\substack{v\in\{1,\ldots,n\} \\ |v-u|>l}} \sum_{\substack{w\in\{1,\ldots,n\} \\ |w-z|>l}} \text{Cov}_{\theta^*}^{\pi}\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n}), \text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right).$$

We use again (A.20) and the Cauchy-Schwartz inequality and observe that

$$= \sum_{v=u-l}^{u+l} \sum_{\substack{w\in\{1,\ldots,n\} \\ |w-z|>l}} \text{Cov}_{\theta^*}^{\pi}\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n}), \text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right)$$

$$\leq \sum_{v=u-l}^{u+l} \sum_{\substack{w\in\{1,\ldots,n\} \\ |w-z|>l}} \left(\mathbb{E}_{\theta^*}^{\pi}\left[\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n})\right)^2\right]\right)^{1/2} \left(\mathbb{E}_{\theta^*}^{\pi}\left[\left(\text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right)^2\right]\right)^{1/2}$$

$$\leq \sum_{v=u-l}^{u+l} \sum_{\substack{w\in\{1,\ldots,n\} \\ |w-z|>l}} \left(\mathbb{E}_{\theta^*}^{\pi}\left[\left(4a_u^0(Z_u)b_v^0(Z_v)\rho^{|v-u|-3}\right)^2\right]\right)^{1/2} \left(\mathbb{E}_{\theta^*}^{\pi}\left[\left(4a_z^0(Z_z)b_w^0(Z_w)\rho^{|w-z|-3}\right)^2\right]\right)^{1/2}$$

$$\leq C \sum_{v=u-l}^{u+l} \left(\mathbb{E}_{\theta^*}^{\pi}\left[\left(a_u^0(Z_u)b_v^0(Z_v)\right)^2\right]\right)^{1/2} \rho^l \sum_{w=1}^{\infty} \left(\mathbb{E}_{\theta^*}^{\pi}\left[\left(a_z^0(Z_z)b_w^0(Z_w)\rho^w\right)^2\right]\right)^{1/2}$$

$$\leq C' \sum_{v=u-l}^{u+l} \left(\mathbb{E}_{\theta^*}^{\pi}\left[\left(a_u^0(Z_u)b_v^0(Z_v)\right)^2\right]\right)^{1/2} \rho^l q_4^{1/2} \sum_{w=1}^{\infty} \rho^w$$

$$\leq C'' q_4(l+1)\rho^l,$$

where $C, C', C'' \in \mathbb{R}$ are constants and q_4 is a bound on the fourth moment of a_u^0 and b_v^0 for all $u, v \in \mathbb{N}$. Similarly, one can show that

$$\sum_{\substack{v\in\{1,\ldots,n\} \\ |v-u|>l}} \sum_{w=z-l}^{z+l} \text{Cov}_{\theta^*}^{\pi}\left(\text{Cov}_{\theta^*}^{\gamma}(a_u, b_v \mid Z_{1:n}), \text{Cov}_{\theta^*}^{\gamma}(a_z, b_w \mid Z_{1:n})\right) = O\left(q_4(l+1)\rho^l\right)$$

Furthermore, note that

$$\sum_{\substack{v\in\{1,\dots,n\}\\|v-u|>l}}\sum_{\substack{w\in\{1,\dots,n\}\\|w-z|>l}}\mathrm{Cov}^{\pi}_{\theta^*}\left(\mathrm{Cov}^{\gamma}_{\theta}(a_u,b_v\mid Z_{1:n}),\mathrm{Cov}^{\gamma}_{\theta}(a_z,b_w\mid Z_{1:n})\right)\le C\rho^{2l}q_4\sum_{v=1}^{\infty}\rho^v\sum_{w=1}^{\infty}\rho^w$$

$$\le C'\rho^{2l}q_4$$

for some constants $C, C' \in \mathbb{R}$. In total we can estimate

$$\mathrm{Cov}^{\pi}_{\theta^*}(\xi_u,\xi_z)-\mathrm{Cov}^{\pi}_{\theta^*}\left(\xi_u^l,\xi_z^l\right)=O\left(q_4(l+1)\rho^l\right) \tag{A.21}$$

Next for $u, v \in \mathbb{N}$, set

$$\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{(u-l):(v+l)})=\mathrm{Cov}^{\gamma}_{\theta^*}\left(a_u(\theta^*)\frac{q^{\gamma,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u}\mid Z_{(u-l):(v+l)})}{p^{\gamma,1:n}_{\theta^*,(u-1):u}(X_{(u-1):u}\mid Z_{(u-l):(v+l)})},\right.$$
$$\left.b_v(\theta^*)\frac{q^{\gamma,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v}\mid Z_{(u-l):(v+l)})}{p^{\gamma,1:n}_{\theta^*,(v-1):v}(X_{(v-1):v}\mid Z_{(u-l):(v+l)})}\mid Z_{(u-l):(v+l)}\right).$$

and

$$\tilde{\xi}_u^l=\sum_{v=u-l}^{u+l}\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{(u-l):(v+l)}).$$

Similarly to (4.9) one can show that

$$\left|\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{1:n})-\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{u-l:v+l})\right|=O\left(a_u^0(Z_u)b_v^0(Z_v)\rho^l\right).$$

It follows that

$$\left|\mathrm{Cov}^{\pi}_{\theta^*}(\xi_u,\xi_z)-\mathrm{Cov}^{\pi}_{\theta^*}\left(\xi_u^l,\xi_z^l\right)\right|$$
$$\le\left|\mathrm{Cov}^{\pi}_{\theta^*}(\xi_u,\xi_z)-\mathrm{Cov}^{\pi}_{\theta^*}\left(\xi_u^l,\xi_z\right)\right|+\left|\mathrm{Cov}^{\pi}_{\theta^*}\left(\xi_u^l,\xi_z\right)-\mathrm{Cov}^{\pi}_{\theta^*}\left(\xi_u^l,\xi_z^l\right)\right|$$
$$=\left|\mathrm{Cov}^{\pi}_{\theta^*}\left(\xi_u-\xi_u^l,\xi_z\right)\right|+\left|\mathrm{Cov}^{\pi}_{\theta^*}\left(\xi_u^l,\xi_z-\xi_z^l\right)\right|$$
$$\le\sum_{v=u-l}^{u+l}\sum_{w=z-l}^{z+l}\left|\mathrm{Cov}^{\pi}_{\theta^*}\left(\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{1:n})-\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{u-l:v+l}),\mathrm{Cov}^{\gamma}_{\theta^*}(a_z,b_w\mid Z_{1:n})\right)\right|$$
$$+\sum_{v=u-l}^{u+l}\sum_{w=z-l}^{z+l}\left|\mathrm{Cov}^{\pi}_{\theta^*}\left(\mathrm{Cov}^{\gamma}_{\theta^*}(a_z,b_w\mid Z_{1:n})-\mathrm{Cov}^{\gamma}_{\theta^*}(a_z,b_w\mid Z_{z-l,v+l}),\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{u-l,v+l})\right)\right|$$
$$\le\sum_{v=u-l}^{u+l}\sum_{w=z-l}^{z+l}\mathbb{E}^{\pi}_{\theta^*}\left[\left(\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{1:n})-\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{u-l:v+l})\right)^2\right]^{1/2}\mathbb{E}^{\pi}_{\theta^*}\left[\left(\mathrm{Cov}^{\gamma}_{\theta}(a_z,b_w\mid Z_{1:n})\right)^2\right]^{1/2}$$
$$+\sum_{v=u-l}^{u+l}\sum_{w=z-l}^{z+l}\left|\mathrm{Cov}^{\pi}_{\theta^*}\left(\mathrm{Cov}^{\gamma}_{\theta^*}(a_z,b_w\mid Z_{1:n})-\mathrm{Cov}^{\gamma}_{\theta^*}(a_z,b_w\mid Z_{u-l:v+l})),\mathrm{Cov}^{\gamma}_{\theta^*}(a_u,b_v\mid Z_{u-l:v+l})\right)\right|$$
$$=O(q_4(l+1)^2\rho^l) \tag{A.22}$$

Again (A.20) leads to

$$\xi_u^l \le a_u^0(Z_u) \sum_{v=u-l}^{u+l} b_v^0(Z_v)\rho^{|v-u|}$$

and further the Hölder inequality implies that for $\kappa = \delta/2$

$$\mathbb{E}_{\theta^*}^\pi\left[(\xi_u^l)^{2+\kappa}\right] = O(q_{4+\delta}l^3),$$

where $q_{4+\delta}$ is an upper bounud on the $(4 + \delta)$th moment of a_u and b_v for all $u, v \in \mathbb{N}$. Finally, we use Theorem 17.2.2 in Ibragimov and Linnik (1971) to bound the covariance of (ξ_u^l, ξ_z^l) by

$$\text{Cov}_\theta^\pi\left(\xi_u^l, \xi_z^l\right) = O\left(q_4 l^3 (\rho^{1/3})^{max(0,|z-u|-4l)}\right) \tag{A.23}$$

Combining (A.21), (A.22) and (A.23), we find that

$$\text{Var}_{\theta^*}^\pi\left(\sum_{u,v=1}^n \text{Cov}_{\theta^*}^\pi(a_u, b_v \mid 1, n)\right)$$

$$= \text{Cov}_{\theta^*}^\pi\left(\sum_{u,v=1}^n \text{Cov}_{\theta^*}^\pi(a_u, b_v \mid 1, n), \sum_{z,w=1}^n \text{Cov}_{\theta^*}^\pi(a_u, b_v \mid 1, n)\right)$$

$$= \sum_{u=1}^n \sum_{z=1}^n \text{Cov}_{\theta^*}^\pi(\xi_u, \xi_z)$$

$$= \sum_{u=1}^n \sum_{z=1}^n O(q_4 l^3 \rho^{\frac{max(0,|u-z|-4l)}{3}}) + O(q_4(l+1)^2\rho^l) + O(q_4(l+1)\rho^l)$$

Choosing $l = \lfloor |u - z|/8 \rfloor$ we obtain that $\text{Var}_{\theta^*}^\pi\left(\sum_{u,v=1}^n \text{Cov}_{\theta^*}^\pi(a_u, b_v \mid 1, n)\right)$ is of order n. $\square$

Proof of Proposition *5.3.* From the definition of α and β it follows that

$$\alpha_i(s) = \sum_{s_{1:i-1}\in S^{i-1}} \nu(s_1)f_{\theta,1}(s_1,z_1)\prod_{j=2}^{i-1}(P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j))P_\theta(s_{i-1},s)f_{\theta,i}(s,z_i)$$

$$= \sum_{k=1}^K \sum_{s_{1:(i-2)}\in S^{i-2}} \nu(s_1)f_{\theta,1}(s_1,z_1)\prod_{j=2}^{i-2}(P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j))$$

$$\times P_\theta(s_{i-2},k)f_{\theta,i}(k,z_i)P_\theta(j,s)f_{\theta,i}(s,z_i)$$

$$= \sum_{j=1}^K \alpha_{i-1}(j)P_\theta(j,s)f_{\theta,i}(s,z_i)$$

and

$$
\begin{aligned}
\beta_{i-1}(s) &= \sum_{s_{i:n}\in S^{n-i+1}} P_\theta(s,s_i)f_{\theta,i}(s_i,z_i)\prod_{j=i+1}^{n}P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j)\\
&= \sum_{j=1}^{K}\sum_{s_{(i+1):n}\in S^{n-i}} P_\theta(s,j)f_{\theta,i}(j,z_i)P_\theta(j,s_{i+1})f_{\theta,i}(s_{i+1},z_{i+1})\prod_{j=i+2}^{n}P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j)\\
&= \sum_{j=1}^{K} P_\theta(s,j)f_{\theta,i}(j,z_i)\sum_{s_{(i+1):n}\in S^{n-i}}P_\theta(j,s_{i+1})f_{\theta,i}(s_{i+1},z_{i+1})\prod_{j=i+2}^{n}P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j)\\
&= \sum_{j=1}^{K}\beta_i(j)P_\theta(s,j)f_{\theta,i}(j,z_i)
\end{aligned}
$$

$\square$

Proof of Proposition *5.10.* Using Proposition 5.4 we find that

$$
\begin{aligned}
&\xi_t(i,j)\\
&= \mathbb{P}_\theta^\nu(X_t=i,X_{t+1}=j\mid Z_1=z_1,\dots,Z_n=z_n)\\
&= \left(\sum_{s_{1:(t-1)}\in S^{t-1}}\sum_{s_{(t+2):n}\in S^{n-t-1}}\nu(s_1)f_{\theta,1}(s_1,z_1)\prod_{j=2}^{t-1}P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j)P_\theta(s_{t-1},i)f_{\theta,t}(i,z_t)\right.\\
&\qquad\left. P_\theta(i,j)f_{\theta,t+1}(j,z_{t+1})P_\theta(j,s_{t+2})f_{\theta,t+2}(s_{t+2},z_{t+2})\prod_{j=t+3}^{n}P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j)\right)/p_\theta^\nu(z_1,\dots,z_n)\\
&= \left(\sum_{s_{1:(t-1)}\in S^{t-1}}\nu(s_1)f_{\theta,1}(s_1,z_1)\prod_{j=2}^{t-1}P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j)P_\theta(s_{t-1},i)f_{\theta,t}(i,z_t)P_\theta(i,j)f_{\theta,t+1}(j,z_{t+1})\right.\\
&\qquad\left. \sum_{s_{(t+2):n}\in S^{n-t-1}}P_\theta(j,s_{t+2})f_{\theta,t+2}(s_{t+2},z_{t+2})\prod_{j=t+3}^{n}P_\theta(s_{j-1},s_j)f_{\theta,j}(s_j,z_j)\right)/p_\theta^\nu(z_1,\dots,z_n)\\
&= \frac{\alpha_t(i)P_\theta(i,j)f_{\theta,t+1}(j,z_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{K}\sum_{j=1}^{K}\alpha_t(i)P_\theta(i,j)f_{\theta,t+1}(j,z_{t+1})\beta_{t+1}(j)}
\end{aligned}
$$

$\square$

Markov chains and Auxiliary results

B.1 A strategy to prove strong consistency of estimators

For maximum likelihood estimation the approach of Wald, see Wald (1949), to prove consistency is straightforward. Here we consider a quasi-likelihood estimator but we see that the approach also works straightforward in this slightly different setting. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $(G, \mathscr{G})$ be a measurable space. Assume that $\Theta \subseteq \mathbb{R}^d$ and let $|\cdot|$ be the d-dimensional Euclidean norm.

Theorem B.1 (Strong consistency). *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of random variables mapping from $(\Omega, \mathscr{F}, \mathbb{P})$ to $(G, \mathscr{G})$. For any $n \in \mathbb{N}$ let $h_n : \Theta \times G^n \to [0, \infty)$ be a measurable function. Assume that there exists an element $\theta^* \in \Theta$ such that for any closed $C \subset \Theta$ with $\theta^* \notin C$ and all $n \in \mathbb{N}$, we have*

$$\limsup_{n\to\infty} \sup_{\theta \in C} \frac{h_n(\theta, W_1, \ldots, W_n)}{h_n(\theta^*, W_1, \ldots, W_n)} = 0 \quad \mathbb{P}\text{-}a.s. \tag{B.1}$$

Let $(\hat{\theta}_n)_{n\in\mathbb{N}}$ be a sequence of random variables mapping from $(\Omega, \mathscr{F}, \mathbb{P})$ to Θ such that

$$\exists c > 0 \ \& \ n_0 \in \mathbb{N} \quad \forall n \geq n_0 : \quad \frac{h_n(\hat{\theta}_n, W_1, \ldots, W_n)}{h_n(\theta^*, W_1, \ldots, W_n)} \geq c, \quad \mathbb{P}\text{-}a.s. \tag{B.2}$$

Then

$$\lim_{n\to\infty} \left|\hat{\theta}_n - \theta^*\right| = 0 \quad \mathbb{P}\text{-}a.s.$$

Proof. For arbitrary $\varepsilon > 0$ define

$$A_\varepsilon^{(1)} := \left\{ \omega \in \Omega \colon \limsup_{n\to\infty} \left|\hat{\theta}_n(\omega) - \theta^*\right| > \varepsilon \right\},$$

$$A_\varepsilon^{(2)} := \left\{ \omega \in \Omega \colon \limsup_{n\to\infty} \sup_{\theta \colon |\theta-\theta^*|\geq\varepsilon} \frac{h_n(\theta, W_1(\omega), \ldots, W_n(\omega))}{h_n(\hat{\theta}_n(\omega), W_1(\omega), \ldots, W_n(\omega))} \geq 1 \right\},$$

$$A_\varepsilon^{(3)} := \left\{ \omega \in \Omega \colon \limsup_{n\to\infty} \sup_{\theta \colon |\theta-\theta^*|\geq\varepsilon} \frac{h_n(\theta, W_1(\omega), \ldots, W_n(\omega))}{h_n(\theta^*, W_1(\omega), \ldots, W_n(\omega))} \geq c \right\}.$$

Note that $A_\varepsilon^{(1)} \subseteq A_\varepsilon^{(2)} \subseteq A_\varepsilon^{(3)}$, where the last inclusion follows by (B.2). Hence, by (B.1) we have

$\mathbb{P}(A_\varepsilon^{(3)}) = 0$ so that

$$\mathbb{P}(A_\varepsilon^{(1)}) = \mathbb{P}\left(\limsup_{n\to\infty} \left|\hat{\theta}_n - \theta^*\right| > \varepsilon\right) = 0,$$

which implies the assertion.   □

The following lemma is useful to verify condition (B.1).

Lemma B.2. *Let $(W_n)_{n\in\mathbb{N}}$ be a sequence of random variables mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(G, \mathcal{G})$ and, as in Theorem B.1, for any $n \in \mathbb{N}$ let $h_n : \Theta \times G^n \to [0, \infty)$ be a measurable function. Assume that there is an element $\theta^* \in \Theta$ such that for any closed $C \subset \Theta$ with $\theta^* \notin C$ we have*

$$\limsup_{n\to\infty} \sup_{\theta\in C} \frac{1}{n} \log h_n(\theta, W_1, \ldots, W_n) < \lim_{n\to\infty} \frac{1}{n} \log h_n(\theta^*, W_1, \ldots, W_n) \quad \mathbb{P}\text{-}a.s. \qquad (B.3)$$

provided that the limit on the right hand-side exists. Then condition (B.1) *is satisfied.*

Proof. Obviously (B.3) implies

$$\log\left(\limsup_{n\to\infty} \sup_{\theta\in C} \left[\frac{h_n(\theta, W_1, \ldots, W_n)}{h_n(\theta^*, W_1, \ldots, W_n)}\right]^{1/n}\right) < 0.$$

This leads to

$$\limsup_{n\to\infty} \sup_{\theta\in C} \left[\frac{h_n(\theta, W_1, \ldots, W_n)}{h_n(\theta^*, W_1, \ldots, W_n)}\right]^{1/n} < 1$$

from which (B.1) follows.   □

B.2   Introduction into Markov Models

In this section we give a short introduction into Markov models. For a detailed survey we refer to Grimmett and Stirzaker (1992). The term 'Markov Model' or 'Markov chain', named after Andrey Markov, originally referred to stochastic models where the probability of a future state only depends on its current state. This property is known as 'Markovian property'. In this section we restrict ourselves to the case where the sample space S is finite and observations are drawn in discrete time. For analogue definitions in general state spaces we refer to Meyn and Tweedie (1992).

Definition B.3. A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with $X_n : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S})$ is a *Markov chain* if it satisfies the *Markov property*, i.e., for all $n, m \in \mathbb{N}$ with $n > m$ and all $x_m, \ldots, x_n \in S$ we have

$$p_{m,n}(x_{n-1}, x_n) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}),$$

where

$$p_{m,n}(x_{n-1}, x_n) := \mathbb{P}(X_n = x_n \mid X_m = x_m, \ldots, X_{n-1} = x_{n-1})$$

The sequence of matrices $(P_n)_{n \in \mathbb{N}}$ generated by

$$P_n(i, j) = p_{n-1,n}(i, j), \qquad n \in \mathbb{N}$$

is called *transition matrices* or *transition kernels*. A Markov chain $(X_n)_{n \in \mathbb{N}}$ is *homogeneous* if

$$P_n = P_{n+1} \quad \forall n > 1.$$

The following theorem provides a connection between the transition matrices at different sample times.

Theorem B.4. *(Chapman-Kolmogorov Equation) Lel $(X_n)_{n \in \mathbb{N}}$ be a Markov chain with transition kernels $(P_n)_{n \in \mathbb{N}}$. For all $n, m \in \mathbb{N}$ with $n > m$ and all $i, j \in S$ it holds that*

$$\mathbb{P}(X_n = j \mid X_m = i) = \left(\prod_{k=m+1}^{n} P_k \right)(i, j) \quad n > m \geq 1, i, j \in S.$$

Proof. The case $n - m = 1$ is trivial. Assume that $n - m > 1$. By Bayes's rule and the Markov property we have

$$\mathbb{P}(X_n = j \mid X_m = i) = \sum_{s_{m+1},\ldots,s_{n-1} \in S} \mathbb{P}(X_{m+1} = s_{m+1}, \ldots, X_{n-1} = s_{n-1}, X_n = j \mid X_m = i)$$

$$= \sum_{s_{m+1},\ldots,s_{n-1} \in S} P_{m+1}(j, s_{m+1}) \prod_{k=m+2}^{n-1} P_{m+1}(s_{k-1}, s_k) P_{m+1}(s_{n-1}, i)$$

$$= \left(\prod_{k=m+1}^{n} P_k \right)(i, j)$$

$\square$

Definition B.5. Let $(X_n)_{n \in \mathbb{N}}$ be a homogeneous Markov chain with transition matrix P. A state $i \in S$ is called *recurrent* if the probability that the Markov chain eventually returns to i is 1, i.e.,

$$\mathbb{P}(X_n = 1 \text{ for some } n > 1 \mid X_1 = i) = 1.$$

If i is not recurrent, it is called *transient*. If all states are recurrent, the Markov chain is called *recurrent*.

For $n \in \mathbb{N}$ with $n > 1$ and $i, j \in S$ let $f_{i,j}(n)$ be the probability of the first passage from i to j, i.e.,

$$f_{i,j}(n) = \mathbb{P}(X_n = j, X_{n-1} \neq j, X_{n-2} \neq j, \ldots, X_2 \neq j \mid X_1 = i)$$

and define

$$f_{i,j} := \sum_{n=2}^{\infty} f_{i,j}(n).$$

We have that j is recurrent if and only if $f_{j,j} = 1$. The following corollary is useful to determine whether a state is recurrent or not.

Corollary B.6. *Let $(X_n)_{n\in\mathbb{N}}$ be a homogeneous Markov chain with transition kernel P and $i, j \in S$. Then it holds that*

i) *State j is recurrent if and only if $\sum_n P^n(j, j) = \infty$ and if this holds $\sum_n P^n(i, i) = \infty$ for all i with $f_{i,j} > 0$.*

ii) *State j is transient if and only if $\sum_n P^n(j, j) < \infty$ and if this holds $\sum_n P^n(i, j) < \infty$ for all i.*

Proof. See page 221 in Grimmett and Stirzaker (1992).   □

Remark B.7. *It follows immediately that $P^n(i, j) \to 0, i, j \in S$ as $n \to \infty$ if j is transient.*

We define $T_j := \min\{n \geq 1 : X_n = j\}$ to be the time of the first visit to j with the convention that $T_j = \infty$ if j is transient and divide the class of recurrent Markov chains into two subclasses.

Definition B.8. Let $(X_n)_{n\in\mathbb{N}}$ be a homogeneous Markov chain with transition kernel P. A recurrent state $i \in S$ is called positive if the *mean recurrence time* defined as

$$\mu_i := \mathrm{E}[T_i | X_1 = i] = \sum_{n=1}^{\infty} f_{i,i}(n) n$$

is finite. Otherwise i is called *null*. Let $d(i) = \gcd\{n \in \mathbb{N} : P^n(i, i) > 0\}$ be the period of i. Here $\gcd(A)$ is the greatest common divisor of A where $A \subset \mathbb{N}^{\mathbb{N}}$. A state $i \in s$ is called *aperiodic* if $d(i) = 1$. A Markov chain is *aperiodic* if all states are aperiodic.

Further for $i, j \in S$ we say *i communicates* with j if there is a positive probability that the chain reaches j starting from i. Then, we write $i \to j$. If also $j \to i$ we say states i and j *intercommunicate* and write $i \leftrightarrow j$. A set $A \in S$ is called *irreducible* if for all $i, j \in A$ we have $i \leftrightarrow j$. A Markov chain is *irreducible* if S is irreducible.

Lemma B.9. *For a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix P at least on state is recurrent.*

Proof. Assume all states are transient. Then by corollary B.6 we have $P^n(i, i) \to \infty$ as $n \to \infty$. This yields to a contradiction since

$$1 = \sum_{j=1}^{K} p_{ij}(0, n) \to 0,$$

as $n \to \infty$.   □

Definition B.10. A distribution $\pi = (\pi^{(1)}, \dots, \pi^{(K)}) \in \mathcal{P}(S)$ is called *invariant distribution* for a Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix P if

$$\pi^{(j)} = \sum_{i \in S} \pi^{(i)} P(i, j), \quad \forall j \in S.$$

Corollary B.11. *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and $i, j \in S$ such that $i \leftrightarrow j$. Then*

   i) i is transient if and only if j is transient

   ii) i is positive recurrent if and only if j is positive recurrent

   iii) i and j have the same period

Proof. We refer to Grimmett and Stirzaker (1992). □

Theorem B.12. *An irreducible Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition matrix P has an invariant distribution $\pi, \pi \in \mathcal{P}(S)$, if and only if all states are positive recurrent. In this case π is given by*

$$\pi^{(i)} = \frac{1}{\mu_i}, \qquad i \in S,$$

where μ_i is the mean recurrence time of state i.

Proof. Sees Grimmett and Stirzaker (1992). □

Remark B.13. *It follows that every irreducible Markov chain with finite state space has an invariant distribution. A homogeneous Markov chain is irreducible if and only if its transition matrix P is irreducible.*

Theorem B.14. *Suppose that P is the transition matrix of an aperiodic, irreducible Markov chain $(X_n)_{n\in N}$ with invariant distribution π and let ρ be an arbitrary distribution on S. Then with probability one it follows that*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{S}} \left|\rho P^n(A) - \pi(A)\right| = 0. \tag{B.4}$$

Proof. We refer to Grimmett and Stirzaker (1992). □

Definition B.15. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and $i \in S$. The sojourn time $S(i)$ of a state i is the number of times steps the Markov chain stays in i, if $X_1 = i$.

Proposition B.16. *(distribution of the sojourn time) Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P and $i \in S$. The sojourn time of state $S(i)$ of i is geometrically distributed with parameter $P(i, i)$.*

Proof. By the Markov property it follows that

$$\begin{aligned}
\mathbb{P}(S(i) = k - 1) &= \mathbb{P}(X_k \neq i, X_{k-1} = i, \ldots, X_2 = i \mid X_1 = i) \\
&= \mathbb{P}(X_k \neq i \mid X_{k-1} = i)\mathbb{P}(X_{k-1} = i, X_{k-2} = i, \ldots, X_2 = i \mid X_1 = i) \\
&= (1 - P(i, i))\mathbb{P}(X_{k-1} = i \mid X_{k-2} = i)\mathbb{P}(X_{k-2} = i, X_{k-3} = i, \ldots, X_2 = i \mid X_1 = i) \\
&= (1 - P(i, i))P(i, i)\mathbb{P}(X_{k-2} = i, X_{k-3} = i, \ldots, X_2 = i \mid X_1 = i) \\
&= (1 - P(i, i))P(i, i)^{k-1},
\end{aligned}$$

where the last line follows from repeating the argument. □

Definition B.17. A sequence of random variables $(X_n)_{n \in \mathbb{N}}$ is *stationary* if the distribution of $X_{n_1}, \ldots, X_{n_k}$ is equal to the distribution of $X_{n_1+r}, \ldots, X_{n_k+r}$ for all $k, r, n_1, \ldots, n_k \in \mathbb{N}$.

Proposition B.18. *A homogeneous Markov chain $(X_n)_{n \in \mathbb{N}}$ with invariant distribution π and transition probability P is a stationary process if $X_1 \sim \pi$.*

Proof. Let $n_1 < n_2 \ldots < n_k$. We use the Markov property and $X_1 \sim \pi$ to obtain

$$
\begin{aligned}
&\mathbb{P}\left(X_{n_1+r} = x_{n_1}, \ldots, X_{n_k+r} = x_{n_k}\right) \\
=&\mathbb{P}\left(X_{n_1+r} = x_{n_1}\right) \mathbb{P}\left(X_{n_2+r} = x_{n_2}, \ldots, X_{n_k+r} = x_{n_k} \mid X_{n_1+r} = x_{n_1}\right) \\
=&\pi(x_{n_1}) \mathbb{P}\left(X_{n_2+r} = x_{n_2} \mid X_{n_1+r} = x_{n_1}\right) \mathbb{P}\left(X_{n_3+r} = x_{n_3}, \ldots, X_{n_k+r} = x_{n_k} \mid X_{n_2+r} = x_{n_2}\right) \\
=&\pi(x_{n_1}) P^{n_2-n_1}(x_{n_1}, x_{n_2}) \mathbb{P}\left(X_{n_3+r} = x_{n_3}, \ldots, X_{n_k+r} = x_{n_k} \mid X_{n_2+r} = x_{n_2}\right) \\
=&\pi(x_{n_1}) \prod_{i=2}^{k} P^{n_i-n_{i-1}}(x_{n_{i-1}}, x_{n_i}) \\
=&\mathbb{P}\left(X_{n_1} = x_{n_1}, \ldots, X_{n_k} = x_{n_k}\right).
\end{aligned}
$$

□

B.3 Auxiliary results

Definition B.19. Let G be a set. A collection of subsets $\mathcal{A} \subset \mathcal{P}(G)$ is a π-system, if $\mathcal{A} \neq \emptyset$ and if $A, B \in \mathcal{A}$ it follows that $A \cap B \in \mathcal{A}$.

Theorem B.20 (Uniqueness theorem for finite measures)**.** *Let $(G, \mathcal{F})$ a measurable space and μ, ν finite measures on $(G, \mathcal{F})$ satisfying $\mu(G) = \nu(G)$. Suppose that for some π-system $\mathcal{A}$ generating $\mathcal{F}$ it holds that $\mu = \nu$ on $\mathcal{A}$. Then $\mu = \nu$ on $\mathcal{F}$.*

Theorem B.21. *Let I be an index set and for every $i \in I$ let $\mathcal{E}_i \subset \mathcal{A}_i$ be a generating system of $\mathcal{A}_i$. Then*

$$\bigotimes_{i \in I} \mathcal{A}_i = \sigma\left(\mathcal{Z}^{\mathcal{E}, \mathcal{R}}\right),$$

where

$$\mathcal{Z}^{\mathcal{E}, \mathcal{R}} = \bigcup_{J \subset I, |J| < \infty} \mathcal{Z}_J^{\mathcal{E}, \mathcal{R}}$$

and $\mathcal{Z}_J^{\mathcal{E}, \mathcal{R}}$ is the set of all rectangular cylinders with basis J.

Proof. See Theorem 14.12 in Klenke (2013). □

Theorem B.22. *Let (G, m) a metric space with its Borel-σ-field $\mathcal{F}$. Two probability measures $\mathbb{Q}_1$ and $\mathbb{Q}_2$ on $\mathcal{F}$ coincide if*

$$\int_G h \mathrm{d}(\mathbb{Q}_1) = \int_G h \mathrm{d}(\mathbb{Q}_2)$$

for all bounded, uniformly continuous function $h : G \rightarrow R$.

Proof. See Theorem 1.2 in Billingsley (1999). □

Definition B.23. Let (G, m) be a metric space with its Borel-σ-field $\mathcal{F}$ and $\mu, \mu_n, n \in \mathbb{N}$ finite measures on $(G, \mathcal{F})$. We say μ_n converges weakly to μ if for any bounded, continuous function $f : G \rightarrow \mathbb{R}$

$$\lim_{n \rightarrow \infty} \int_G f \mathrm{d}(\mu_n) = \int_G f \mathrm{d}(\mu)$$

holds.

Lemma B.24. *Let W be a non-central chi-squared distributed random variable with degree of freedom 1 and non-centrality parameter λ. The moment-generating function $m(t)$ is given by*

$$m(t) = \mathbb{E}\left[\exp(tW)\right] = (1 - 2t)^{-1/2} \exp\left(\frac{\lambda t}{1 - 2t}\right)$$

Lemma B.25. *For $d \in \mathbb{N}$ let $A \in \mathbb{R}^{d \times d}$ be a symmetric and positive definite matrix with eigenvalues $\lambda_i, i = 1, \ldots d$. Then*

(i) The eigenvalues of the inverse matrix A^{-1} are given by $\lambda_i^{-1}, i = 1, \ldots d$.

(ii) The eigenvalues of $A^T A$ are given by $\lambda_i^2, i = 1, \ldots d$.

(iii) There exists a symmetric, positive definite matrix $A^{1/2}$ such that

$$A^{1/2} A^{1/2} = A.$$

(iv) The inverse matrix A^{-1} is symmetric and positive definite.

Proof. For $i \in \mathbb{N}$ let v_i an eigenvector of λ_i. It follows that
(i)

$$A v_i = \lambda_i v_i \Rightarrow A^{-1} A v_i = A^{-1} \lambda_i v_i \Rightarrow \lambda_i^{-1} v_i = A^{-1} v_i,$$

(ii)

$$A^T A v_i = A^T \lambda_i v_i = \lambda_i A v_i = \lambda_i^2 v_i.$$

(iii) Since A is symmetric there exists an orthogonal matrix $U \in \mathbb{R}^{d \times d}$ such that

$$A = UDU^T,$$

where D is a diagonal matrix having the eigenvalues on the diagonal. Let $D^{1/2} \in \mathbb{R}^{d \times d}$ be the diagonal matrix with diagonal $(\lambda_1^{1/2}, \ldots, \lambda_d^{1/2})^T$. It holds that

$$A = UDU^T = UD^{1/2}D^{1/2}U^T = UD^{1/2}U^T UD^{1/2}U^T.$$

Set $A^{1/2} = U D^{1/2} U^T$.

(iv) Let $D^{-1} \in \mathbb{R}^{d \times d}$ denote the diagonal matrix with diagonal $(\lambda_1^{-1}, \dots, \lambda_d^{-1})$. Observe that

$$U D^{-1} U^T U D U^T = I_q.$$

It follows that $A^{-1} = U D^{-1} U^T$.   □

Bibliography

P. Acred, D. M. Brown, D. H. Turner, and M. J. Wilson. Pharmacology and chemotherapy of ampicillin—a new broad-spectrum penicillin. *British Journal of Pharmacology and Chemotherapy*, 18(2):356–369, Apr. 1962.

P. Ailliot and F. Pene. Consistency of the maximum likelihood estimate for non-homogeneous Markov-switching models. *arXiv/1306.2116*, 2013.

F. G. Ball and J. A. Rice. Stochastic models for ion channels: Introduction and bibliography. *Mathematical Biosciences*, 112(2):189–206, Dec. 1992.

A. R. Barron. The Strong Ergodic Theorem for Densities: Generalized Shannon-McMillan-Breiman Theorem. *The Annals of Probability*, 13(4):1292–1303, Nov. 1985.

A. Bartsch, F. Pein, C. Kattner, S. Llabres, M. Schön, M. Diehn, U. Zachariae, M. Tanabe, A. Munk, and C. Steinem. Time-resolved interaction of single ampicillin molecules with Neisseria meningitidis PorB. *submitted*, 2017.

M. Basseville and A. Benveniste. Desgin and comparative study of some sequential jump detection algorithms for digital signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(3):521–535, June 1983.

L. E. Baum and J. A. Eagon. An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bulletin of the American Mathematical Society*, 73(3):360–363, 1967.

L. E. Baum and T. Petrie. Statistical Inference For Probabilistic Functions Of Finite State Markov Chains. *Ann. Math. Statist.*, 37(6):1554–1563, 1966.

L. E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41:164–171, 1970.

P. J. Bickel, Y. Ritov, and T. Rydén. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *The Annals of Statistics*, 26(4):1614–1635, Aug. 1998.

P. Billingsley. *Convergence of Probability Measures*. Wiley-Interscience, New York, 2 edition, July 1999.

O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer, New York ; London, 1st ed. 2005. corr. 2nd printing 2007 edition, Sept. 2007.

K. S. Cole. Dynamic electrical characteristics of the squid axon membrane. *Archives des sciences physiologiques*, 3(2):253–258, 1949.

D. Colquhoun and A. G. Hawkes. A Note on Correlations in Single Ion Channel Records. *Proceedings of the Royal Society of London B: Biological Sciences*, 230(1258):15–52, Feb. 1987.

M. de Gunst and J. Schouten. Model selection and parameter estimation for ion channel recordings with an application to the K+ outward-rectifier in barley leaf. *J Math Biol*, 50(3):233–56, 2005.

D. del Camino, M. Holmgren, Y. Liu, and G. Yellen. Blocker protection in the pore of a voltage-gated K+ channel and its structural implications. *Nature*, 403(6767):321–325, Jan. 2000.

A. H. Delcour. Solute uptake through general porins. *Frontiers in Bioscience: A Journal and Virtual Library*, 8:d1055–1071, May 2003.

A. H. Delcour. Outer membrane permeability and antibiotic resistance. *Biochimica Et Biophysica Acta*, 1794(5):808–816, May 2009.

S. D. Demo and G. Yellen. Ion effects on gating of the Ca(2+)-activated K+ channel correlate with occupancy of the pore. *Biophysical Journal*, 61(3):639–648, Mar. 1992.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood From Incomplete Data Via the EM Algorithm (C/R: P22-37). *Journal of the Royal Statistical Society, Series B, Methodological*, 39:1–22, 1977.

M. Devices. *The Axon Guide: A Guide to Electrophysiology and Biophysics Laboratory Techniques*. 2008.

R. Douc and C. Matias. Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli*, 7(3):381–420, June 2001.

R. Douc, E. Moulines, and T. Ryden. Asymptotic Properties of the Maximum Likelihood Estimator in Autoregressive Models with Markov Regime. *The Annals of Statistics*, 32(5): 2254–2304, Oct. 2004.

R. Douc, E. Moulines, J. Olsson, and R. van Handel. Consistency of the maximum likelihood estimator for general hidden Markov models. *The Annals of Statistics*, 39(1):474–513, Feb. 2011.

R. J. Elliott, L. Aggoun, and J. B. Moore. *Hidden Markov Models: Estimation and Control*. Springer Science & Business Media, Sept. 2008. Google-Books-ID: WQzlBwAAQBAJ.

V. Genon-Catalot and C. Laredo. Leroux's method for general hidden Markov models. *Stochastic Processes and their Applications*, 116(2):222–243, Feb. 2006.

M. Gerber, T. Kaufmann, and B. Pfister. Extended Viterbi algorithm for optimized word HMMS. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4932–4935, May 2011.

F. L. Gland and L. Mevel. Basic Properties of the Projective Product with Application to Products of Column-Allowable Nonnegative Matrices. *Mathematics of Control, Signals and Systems*, 13(1):41–62, Feb. 2000a.

F. L. Gland and L. Mevel. Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models. *Mathematics of Control, Signals and Systems*, 13(1):63–93, Feb. 2000b.

R. M. Gray. *Probability, Random Processes, and Ergodic Properties*. Springer Publishing Company, Incorporated, 2nd edition, 2009.

R. M. Gray and J. C. Kieffer. Asymptotically Mean Stationary Measures. *The Annals of Probability*, 8(5):962–973, Oct. 1980.

G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 1992.

O. P. Hamill, A. Marty, E. Neher, B. Sakmann, and F. J. Sigworth. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflugers Archiv: European Journal of Physiology*, 391(2):85–100, Aug. 1981.

B. Hille. *Ionic Channels of Excitable Membranes*. Sinauer, Sunderland, Mass, 3rd ed. 2001 edition, Aug. 2001.

T. Hotz, O. M. Schütte, H. Sieling, T. Polupanow, U. Diederichsen, C. Steinem, and A. Munk. Idealizing Ion Channel Recordings by a Jump Segmentation Multiresolution Filter. *IEEE Transactions on NanoBioscience*, 12(4):376–386, Dec. 2013.

R. Hsiao, Y. C. Tam, and T. Schultz. Generalized Baum-Welch algorithm for discriminative training on large vocabulary continuous speech recognition system. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3769–3772, Apr. 2009.

I. A. Ibragimov and I. I. V. Linnik. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff., 1971.

J. L. Jensen. Asymptotic normality of M-estimators in nonhomogeneous hidden Markov models. *Journal of Applied Probability*, 48A:295–306, Aug. 2011a.

J. L. Jensen. Central limit theorem for functions of weakly dependent vriables. In *Proceedings of the 2011 World Statistics Congress*, page 5362, Aug. 2011b.

A. Klenke. *Probability Theory: A Comprehensive Course*. Springer, New York, 2nd edition, Sept. 2013.

B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stochastic Processes and their Applications*, 40(1):127–143, 1992.

T. A. Louis. Finding the Observed Information Matrix When Using the EM Algorithm. *Journal of the Royal Statistical Society, Series B: Methodological*, 44:226–233, 1982.

S. P. Meyn and R. L. Tweedie. Stability of Markovian Processes I: Criteria for Discrete-Time Chains. *Advances in Applied Probability*, 24(3):542–574, 1992.

M. Olesky, M. Hobbs, and R. A. Nicholas. Identification and analysis of amino acid mutations in porin IB that mediate intermediate-level resistance to penicillin and tetracycline in Neisseria gonorrhoeae. *Antimicrobial Agents and Chemotherapy*, 46(9):2811–2820, Sept. 2002.

B. A. Oppenheim. Antibiotic resistance in Neisseria meningitidis. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 24 Suppl 1:S98–101, Jan. 1997.

F. Pein, T. Hotz, H. Sieling, and T. Aspelmeier. stepR: Multiscale Change-Point Inference, May 2017a.

F. Pein, I. Tecuapetla-Gómez, O. M. S. C. Steinem, and A. Munk. Fully-Automatic Multiresolution Idealization for Filtered Ion Channel Recordings: Flickering Event Detection. *arXiv:1706.03671 [stat]*, June 2017b. arXiv: 1706.03671.

D. Pouzo, Z. Psaradakis, and M. Sola. Maximum Likelihood Estimation in Possibly Misspecified Dynamic Models with Time-Inhomogeneous Markov Regimes. *arXiv:1612.04932 [math, stat]*, Dec. 2016. arXiv: 1612.04932.

F. Qin, A. Auerbach, and F. Sachs. Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophysical Journal*, 79(4):1928–1944, Oct. 2000.

O. W. Rechard. Invariant measures for many-one transformations. *Duke Mathematical Journal*, 23(3):477–488, Sept. 1956.

T. Ryden. EM versus Markov chain Monte Carlo for estimation of hidden Markov models: a computational perspective. *Bayesian Analysis*, 3(4):659–688, Dec. 2008.

B. Sakmann and E. Neher. Patch clamp techniques for studying ionic channels in excitable membranes. *Annual Review of Physiology*, 46:455–472, 1984.

B. Sakmann and E. Neher. *Single-Channel Recording*. Springer, New York, NY, 2nd ed. 1995 edition, June 2010.

I. Siekmann, L. E. Wagner, D. Yule, C. Fox, D. Bryant, E. J. Crampin, and J. Sneyd. MCMC Estimation of Markov Models for Ion Channels. *Biophysical Journal*, 100(8):1919–1929, Apr. 2011.

F. J. Sigworth. Open channel noise. II. A test for coupling between current fluctuations and conformational transitions in the acetylcholine receptor. *Biophysical Journal*, 49(5):1041–1046, May 1986.

M. Tanabe, C. M. Nimigean, and T. M. Iverson. Structural basis for solute transport, nucleotide regulation, and immunological recognition of Neisseria meningitidis PorB. *Proceedings of the National Academy of Sciences of the United States of America*, 107(15):6811–6816, Apr. 2010.

H. T. Tien and A. L. Ottova. The lipid bilayer concept and its experimental realization: from soap bubbles, kitchen sink, to bilayer lipid membranes. *Journal of Membrane Science*, 189(1): 83–117, July 2001.

D. J. Triggle. *Voltage-Gated Ion Channels as Drug Targets*. Wiley-VCH Verlag GmbH, Apr. 2006.

L. Venkataramanan, R. Kuc, and F. J. Sigworth. Identification of hidden Markov models for ion channel currents .III. Bandlimited, sampled data. *IEEE Transactions on Signal Processing*, 48 (2):376–385, Feb. 2000.

A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, Apr. 1967.

A. Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20(4):595–601, Dec. 1949.

M. Winterhalter. Black lipid membranes. *Current Opinion in Colloid & Interface Science*, 5(3): 250–255, July 2000.

C. F. J. Wu. On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, 11(1): 95–103, Mar. 1983.

G. Yellen. Single Ca2+-activated nonselective cation channels in neuroblastoma. *Nature*, 296 (5855):357–359, Mar. 1982.

G. Yellen. The moving parts of voltage-gated ion channels. *Quarterly Reviews of Biophysics*, 31 (3):239–295, Aug. 1998.

W. Zucchini and I. L. Macdonald. *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman & Hall/CRC, Boca Raton, 2rev ed edition, Apr. 2009.

Manuel Diehn
Curriculum Vitae

PERSONAL DETAILS

Birth	April 17, 1988
Address	Ilmenauer Weg 5, Göttingen
Mail	`mdiehn1@gwdg.de`

EDUCATION

Research Assistant and Ph.D. Student in Mathematics 10/2013-present

Georg-August-University of Göttingen

MSc. in Business Mathematics 10/2011-10/2013

Georg-August-University of Göttingen

BSc. in Mathematics 10/2008-10/2011

Georg-August-University of Göttingen

WORK EXPERIENCE

Internship 08/2012-10/12

Deloitte Deutschland, FRS, Düsseldorf

Development and Implementation of a program for CVA calculation. BoB-membership.

Internship 02/2012-04/2012

d-fine, Zürich

Consultant in the IT department of a Swiss bank, Development and Implementation of a testing procedure for the trading system Front Arena.

Internship 03/2011-05/2011

ERGO Versicherung , Actuarial Department, Hamburg

Care and maintenance of the internal computing program, Analysis of the effects of the increase of the internal calculation interest rate.

SKILLS

Languages	German (mother tongue), English (fluent)
Software	Matlab, LaTeX, R, C++